SURVIVAL TIME AND PREDICTORS OF MORTALITY AMONG PATIENTS UNDER MULTI-DRUG RESISTANT TUBERCULOSIS TREATMENT IN GONDAR UNIVERSITY HOSPITAL

BY

ASAMENEW ENDAWEKE WALE

THESIS SUBMITTED TO

THE DEPARTMENT OF STATISTICS

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN BIOSTATITICS

UNIVERSITY OF GONDAR, GONDAR, ETHIOPIA

MARCH 2015

SURVIVAL TIME AND PREDICTORS OF MORTALITY AMONG PATIENTS UNDER MULTI-DRUG RESISTANT TUBERCULOSIS TREATMENT IN GONDAR UNIVERSITY HOSPITAL

BY

ASAMENEW ENDAWEKE WALE

ADVISOR

SALIE AYALE (PhD)

UNIVERSITY OF GONDAR, GONDAR, ETHIOPIA

MARCH 2015

DECLARATION

I, the undersigned, declare that the thesis is my original work, has not been presented for Degrees in any other University and all sources of materials used for the thesis have been duly acknowledged.

Declared by:

Name: Asamenew Endaweke Wale

Signature: ----------

Date: ----------

APPROVAL SHEET- 1 -

This is to certify that the thesis entitled "**SURVIVAL TIME AND PREDICTORS OF MORTALITY AMONG PATIENTS UNDER MULTI-DRUG RESISTANT TUBERCULOSIS TREATMENT IN GONDAR UNIVERSITY HOSPITAL"**, Submitted in partial fulfillment of the requirements for the degree of Master of Science in Statistics with a specialization of Biostatistics of the Graduate Program of the Department of statistics, University of Gondar, and is a record of original research carried out by **Asamenew Endaweke Wale, I.D.No 5103/05** under my supervision, and no part of the thesis has been submitted for any other degree or diploma.

The assistance and the help received during the course of this investigation have been duly acknowledged. Therefore, I recommend that it may be accepted as fulfilling the thesis requirements.

_____ _____ _____

Name of Advisor Signature Date

APPROVAL SHEET -2

We, the undersigned, members of the Board of Examiners of the final open defense by **Asamenew Endaweke Wale** have read and evaluated his thesis entitled "**SURVIVAL TIME AND PREDICTORS OF MORTALITY AMONG PATIENTS UNDER MULTI-DRUG RESISTANT TUBERCULOSIS TREATMENT IN GONDAR UNIVERSITY HOSPITAL:** A **CASE OF STUDY GONDAR UNIVERSITY HOSPITAL",** and examined the candidate. This is therefore to certify that the thesis has been accepted in partial fulfillment of the requirements for the degree of Master of Science in Statistics with specialization of Biostatistics Statistics.

_____	_____	_____
Name of Chair Person	Signature	Date
_____	_____	_____
Name of Advisor	Signature	Date
_____	_____	_____
Name of Internal Examiner	Signature	Date
_____	_____	_____
Name of External Examiner	Signature	Date

Table of Contents

List of Tables

List of Figures

Acronyms

AFB	acid-fast bacilli
AIC	Akaki Information Criteria
AIDS	Acquired Immune Deficiency Syndrome
BIC	Bayesian Information Criteria
CI	Confidence Interval
NDACA	National Drug Administration and Control Authority
DOT	directly observed therapy
DST	drug susceptibility testing
EFMOH	Ethiopian Federal ministry of health
GLC	Green Light Committee
HIV	Human Immunodeficiency Virus
HR	Hazard Ratio
INH	isoniazid
KM	Kaplan-Meier
LR	Likelihood Ratio
MDR-TB	Multidrug resistant tuberculosis
MLE	Maximum Likelihood Estimator
MOH	Ministry Of Health
RIF	Rifampicin
SE	Standard Error
SPSS	Statistical Package for Social Science
TB	Tuberculosis
UNITAID	United nation fund to fight AIDS
US	united States
WHO	World health organization
XDR-TB	Extensively drug resistant tuberculosis

Acknowledgments

First, and foremost, I thank God for giving me the opportunity to pursue my graduate study in the Department of Statistics, University of Gondar.

I would like to sincerely thank Dr. Salie Ayalew, my advisor, for his guidance and encouragement from the beginning to the end of the study.

I would like to thank University of Gondar referral Hospital MDR TB staffs who allowed and facilitate all the inputs for data collection.

I am highly indebted to all my families who were the source of strength towards the successful completion of the study.

I acknowledge with appreciation the help rendered by all my dear friends who encouraged me to work hard.

Lastly, I would like to thank Dire Dawa University for sponsoring me to attend this program in general.

ABSTRACT

Multi drug resistant Tuberculosis (MDR TB) is a chronic infection disease that has a major health problem over the centuries and it has accounted for more human misery suffering and loss of earning and failure of economic and social development than any other disease. The aim of this study was to identifying the risk factors associated with time to death among MDR TB patients treated under directly observed short course treatment program in Health facilities at Gondar University hospital, Ethiopia. The data for this study was obtained by accessing medical records of MDR TB patients registered during August 2010 to August 2014and treated in Gondar University Hospital. Kaplan-Meier estimation method, Cox proportional hazard regression model and parametric regression models were applied. The Cox proportional hazard analysis indicated that therapeutic delay after one month, number of drugs taken only INH and RIF, alcohol use and HIV positive co-infected patients are significantly contribute to shorter survival time whereas patients without any clinical complication and with no chronic co-infection were significantly contributed to longer survival time of MDR TB patients. The result from Weibull regression analysis showed that elder MDR TB patients, therapeutic delay after one month, MDR TB patients taken only INH and RIF at initiation, patients who took alcohol, MDR TB patients with different clinical complication, previously not treated MDR TB patients, HIV positive patients and MDR TB patients with different chronic co-infections significantly contribute a shorter survival time of MDR TB patients. The major factors that affect the survival of MDR-TB patients are Age, therapeutic delay, number of drugs taken at initiation, alcohol use, any clinical complication, MDR category, HIV co-infection and chronic co-infection.

Key Words: *MDR TB, Kaplan-Meier, Cox Proportional Hazard, Weibull Regression Model, Survival Analysis, Ethiopia*

CHAPTER ONE

1. INTRODUCTION

1.1 Background of the study

Multi-drug resistant tuberculosis (MDR-TB) is defined as resistant to at least isoniazid (INH) and Rifampicin (RIF), the two most important first line anti-Tb drugs. MDR-TB is a man-made problem. When patients stop taking or do not take enough of the right medications, the poor quality of anti-tuberculosis drugs supply and due to the poor follow up of the means of tuberculosis transmission leads to the emergence and spread of mycobacterium tuberculosis strains resistant to multiple drugs which represent a major public health problem in a number of countries and an obstacle to effective global TB control (WHO, 2011).

MDR-TB is a droplet infection and is easily transmitted to immune compromised individuals, especially to the HIV infected. Experiences in African countries showed that transmission among HIV infected individuals results in MDR-TB micro-epidemics, both non societal and societal. Multidrug resistance can be primary or acquired. Primary resistance is defined as resistance in patients without a history or other evidence of previous anti-tuberculosis drug treatment. Acquired drug resistance occurs in those who have previously received anti-tuberculosis treatment for at least one month and in those with treatment failures and relapses. The prevalence of primary multidrug resistance is 3% and that of acquired multidrug resistance is 12% in India (WHO, 2008). Although drug resistance started appearing in patients with M. tuberculosis infection soon after the introduction of effective anti-tuberculosis drugs, multidrug-resistant tuberculosis was not a major public health problem until the early 1990s when HIV infection became a global epidemic. A history of tuberculosis and previous anti-tuberculosis treatment are the most important factors identified in the causation of multidrug-resistant tuberculosis. Factors which predispose a patient to development of multidrug-resistant tuberculosis include (Yashodhara B. *et al*, 2010):

- Incomplete treatment
- Inadequate treatment
- Errors in tuberculosis management, such as use of a single anti-tuberculosis drug
- Addition of a single drug to a failing regimen
- Failure to identify pre-existing resistance

- Initiation of an inadequate primary regimen
- Failure to identify and address no adherence to treatment
- Noncompliance
- Inappropriate isoniazid preventive therapy
- Variations in bioavailability of anti-tuberculosis drugs

MDR-TB cases are classified into two categories. The one who has primary resistance, a newly registered episode of TB in a patient who, in response to direct questioning, denies having had any prior anti-TB treatment (for less than one month), and in countries where adequate documentation is available, for whom there is no evidence of such history and those who have acquired resistance, A newly registered episode of TB in a patient who, in response to direct questioning admits having been treated for TB for one month or more, or, in countries where adequate documentation is available, there is evidence of such history. Chemoprophylaxis should not be considered treatment for TB consisting of the majority of cases. Insufficient previous treatment is a strong prognostic factor in the development of MDR-TB. Many of the MDR-TB patients had been taking anti-TB drugs for a long time, and often irregularly, which resulted in treatment failure (Espinal M. *et al,* 2001 and Seyed M. *et al,* 2005).

MDR-TB patients respond poorly to short course chemotherapy and need to be treated intensively for up to 24 months with a regimen based on reserve anti-tuberculosis drugs (WHO, 2003). MDR-TB is an increasing global problem. The extent and burden of MDR-TB varies significantly from country to country. Globally, the proportion is higher in patients who had previously received anti-tuberculosis (anti-TB) treatment reflecting the failure of programs designed to ensure complete cure of patients with tuberculosis (Baptista I. *et al,* 2008).

WHO estimates that, 440 000 people had MDR-TB in 2010 and that result a 150 000 deaths from MDR-TB worldwide. In 2010, the largest WHO MDR-TB survey reported the highest rates ever of MDR-TB, with peaks of up to 28% of new TB cases in some settings of the former Soviet Union, including regions sharing borders with the European Union (WHO, 2011). World Health Organization estimated that there were about 0.5 million new MDR-TB cases in the world in 2011. About 60% of these cases occurred in Brazil, China, India, the Russian Federation and South Africa alone. Asia bears the burden of the epidemic as almost 50% of MDR-TB cases worldwide are estimated to occur in China and India .High prevalence of MDR-TB was also

found among new cases in Ecuador, and Israel. Central Europe and Africa, in contrast, reported the lowest median levels of drug resistance (WHO, 2013).

A total of 94 000 TB cases eligible for MDR-TB treatment (84 000 with MDR-TB and 10 000 with rifampicin resistance detected using Expert MTB/ RIF) were notified in 2012, mostly by European countries, India and South Africa. This represented progress compared with 2011, when 62 000 MDR-TB cases and 4000 rifampicin resistant TB cases were detected; the largest increases between 2011 and 2012 were in India, South Africa and Ukraine (WHO, 2013). However, worldwide and in most countries with a high burden of MDR-TB, less than one third of the TB patients estimated to have MDR-TB were actually detected in 2012. Just over 77 000 people with MDR-TB were started on second line treatment in 2012, equivalent to 82% of the 94 000 newly detected cases that were eligible for such treatment globally. Report and published in 2004 have MDR-TB rates >2.0% of all combined TB cases in Africa. This finding suggests that completing DRSs for all or most countries in the AFRO region is urgently needed and that the MDR-TB threat in Africa could be much higher than originally assessed by WHO in its previous report in 2004. Drug-resistant strains, along with HIV/AIDS, are causing the biggest challenge to efficient management and control of TB. The lower rates of MDR-TB in Africa, when compared with rates in Eastern Europe or South America, could be related to the fact that for many years Africa was neglected and TB was not treated. Alternatively, later introduction of rifampin in drug regimens is often cited as an explanation for these low rates. Meanwhile in Africa, where little data is available, an estimated 69 000 cases emerged in 2009, but the vast majority of them went un-diagnosed (WHO, 2013 and EFMOH, 2013).

According to WHO 2008, in Ethiopia, 5825 MDR-TB cases (4964 among newly diagnosed and 861 among previously treated TB cases) were estimated to have occurred in 2006. According to the anti-TB drug resistance survey conducted nationwide in 2005 (EHNRI/EFMOH), among 804 newly diagnosed TB cases 1.6% were found to be infected with MDR-TB. The rate of MDR TB among specimens from 76 previously treated TB cases was 11.8%. The same survey reported that, TB with Isoniazid mono-resistance and Rifampicin mono-resistance, among new TB cases, was 2% and 1%, respectively. Notified prevalence of mono-resistance to INH and Rifampicin among previously treated TB cases was 5.3% and 1.3%, respectively. Based on the prevalence rate from the survey and TB case notification in 2007/08 the magnitude of MDR-TB in Ethiopia

was estimated to be 997 cases, which includes 651 and 346 MDR-TB cases among newly diagnosed and re-treatment cases respectively (EFMOH, 2009).

The treatment of MDR-TB with second line drugs is long, complex and costly, and has a considerable rate of adverse effects. Recent progress has been made globally in improving policy environment and advances in TB drug development to embark on trials of MDR-TB treatment to identify optimal treatment protocol for drug resistant Tuberculosis. The TB Control Program in Ethiopia has not yet started managing MDR-TB cases with efficient second line protocols. No second line anti-TB drugs are available in Ethiopia, with the exception of fluoroquinolone, and the importation of any drug is as per the regulation of the National Drug Administration and Control Authority (DACA). But there are evidences that a small cohort of patients were treated in some facilities in Addis Ababa and regions, The Ethiopian Government has identified MDR-TB as one of priority public health problem and it is committed to initiate comprehensive treatment for MDR-TB cases in the country. The EFMOH has also clearly endorsed the mechanism of single procurement and controlled use of the second line anti-TB drugs, when they are available, after acceptance from the Green Light Committee (GLC) and stands for this unique channel. Management of MDR-TB will be an integral component of the NTP and will be implemented through the existing health care delivery system (EFMOH, 2013).

The FMOH has set up a specific group of specialists to provide expert advices on MDR-TB issues, and has drafted a Plan of Action for MDR-TB control and care. The MDR-TB Technical Working Group was established by the EFMOH and other partners to support MDR-TB activities in the country. It was the decision of the EFMOH to develop clinical and program management guideline for DR-TB with the technical support of MDR-TB Technical Working Group and technical assistance of well renowned international experts in the area, before launching case management in public institutions under close supervision and follow up of National TB control program. In 2002, MDR-TB in Ethiopia was also reported in about 1.2% of new cases and 12% of re- treatment cases. According to WHO 2009 report there are 1.6% MDR-TB patients among all new TB cases and 12% MDR-TB patients among previously treated TB cases in Ethiopia. Of the 27 countries with a high burden of MDR-TB and extensively drug resistant TB (XDR-TB), 13 countries with data on treatment outcomes for MDR-TB cases reported a success by 25%-82% among patients that started on treatment in 2007 (WHO, 2009).

1.2 Statement of the problem

World health organization estimates that there were about 450,000 new (incident) MDR-TB cases in the world in 2012. Since the treatment of MDR-TB in Gondar university hospital started recently, the survival rate and its determinant factors among patients under MDR-TB treatment are not described. Randomized controlled trials and other researches are done in moderate evidence and methodology for optimizing treatment regimens, accessing the risk factors and estimating survival times in MDR-TB patients. In addition treatment of pediatric MDR-TB and survival time of MDR-TB patients including treatment duration exists in current condition in Gondar university hospital. It is believed that, in resource poor countries like Ethiopia the survival of MDR-TB patients under DOTS program depends on a variety of factors, which may also vary greatly with socio-demographic, behavioral risk and health factors. So, Identification of survival time and risk factors of mortality in MDR-TB cases is essential for proper planning and effective implementation of MDR-TB treatment. Therefore it is important to examine the risk factors for the survival of MDR-TB patients and clearly identifying the risk factor of mortality of MDR-TB patients treated under University of Gondar hospital. The crucial questions that the study answers were:

- What are the major factors that facilitate the death of MDR-TB patients under treatment?
- What are the survival probabilities of MDR-TB patients at time t?
- What is the median survival time of MDR- TB patients under the treatment program?
- Which groups among various levels of factors have high risk of death?

1.3 Objectives of the study

1.3.1 General objective

The general objective of this study is to estimate the survival time and predictors of mortality among patients under Multi-Drug Resistant Tuberculosis treatment at Gondar University Hospital.

1.3.2 Specific objectives

The specific objectives of the study are:

- ✓ Estimating the survival time of MDR-TB patients.

- ✓ To compare the risk between different exposure groups and associated factors with survival time of patients under MDR-TB treatment.
- ✓ To develop a statistical model that predicts the survival of MDR-TB patients in the hospital using significant risk factors.

1.4 Significance of the study

Expanding access to MDR-TB therapy is urgently needed, yet poor implementation of such therapy can worsen the problem of XDR-TB. Understanding risk factors for poor treatment outcomes (death) among MDR-TB patients is necessary to improve treatment outcomes. Therefore, examining a cohort who received a standardized second-line therapy and management of MDR-TB to determine the overall survival time and its determinant factors has a great importance. In addition understanding risk factors for death among MDR-TB patients is necessary to improve treatment outcomes. Since the treatment of MDR-TB in Ethiopia started recently, the survival time and its determinant factors among patients under MDR-TB treatment are not well described. Therefore, identification of survival time and risk factors of mortality in MDR-TB cases is essential for proper planning and effective implementation of MDR-TB treatment. Estimating the survival time and identifying predictors of mortality for MDR-TB patients is also important to inform public health authorities, policy makers and stakeholders about risk factors for the survival time and mortality of MDR-TB patients. The findings would help to bring MDR-TB problem to the agenda of public health policy makers, researchers, and the public at large, so that appropriate treatment and control strategies are implemented along with a population wide surveillance intervention.

1.5 Limitation of the Study

- ➢ The data were extracted from medical records of those who already visited and registered at the respective health facilities; it may be subjected to bias.
- ➢ The study was restricted to adults, and results might not be applicable to infants and children.
- ➢ The study is based on only the set of data for which complete information on survival times are available because of missing values.
- ➢ Underestimation of mortality due to lost to follow up patients included in the study. In addition to this all deaths are assumed to be caused by MDR-TB.

CHAPTER TWO

2. LITERATURE REVIEW

2.1 Diagnosis of multi drug resistant tuberculosis

The first step in diagnosing drug-resistant TB is to recognize that the patient is at risk and to expedite the laboratory diagnosis of TB. The diagnosis of tuberculosis (TB) frequently requires a high index of suspicion, especially in low-prevalence areas. Once TB is considered, sputum or other specimens for acid-fast bacilli (AFB) smear, growth detection, and susceptibility testing are collected. The possibility of drug-resistant TB should be considered simultaneously with specimen collection and selection of the initial treatment regimen. Failure to consider the possibility of drug-resistant TB until drug susceptibility tests return weeks to months later can result in unnecessarily inadequate drug regimens.

While a number of diagnostic tests exist, TB is typically diagnosed in many settings throughout the world based on clinical symptoms and on a sputum smear test in which the sputum smear is stained and viewed under a microscope. The results are often confirmed by culturing the sputum to check for growth of the TB bacteria. However, the diagnosis of MDR-TB is more difficult as it requires special laboratory diagnostic capacity for drug-susceptibility testing, but many countries, especially in low-income settings, have little or no diagnostic capacity for MDR-TB. Globally, only 5% of new cases and 9% of previously treated cases received testing for drug-resistance. Detection of TB without testing for drug-resistance can lead to poor treatment outcomes, additional and unnecessary suffering and costs for patients, as well as further spread of drug resistant strains.

2.2 Treatment of Multi Drug Resistant Tuberculosis

In order to bring new treatments to deal in increased infrastructure, new trial designs, increased clinical trial capacity, clear regulatory guidelines, and biomarkers for prediction of long-term outcome and ultimately novel drug combinations are needed. In the last 40 years there has been little advancement in TB treatment, which has contributed to a rise in multidrug-resistance. There are currently several drugs available for the treatment of TB. However, these treatments are older and a significant amount of drug resistance has developed over the years. When first-

line drugs fail, second-line drugs are used to treat MDR-TB. They must be taken for up to two years in order to cure the infection, placing patients at an increased risk for side-effects and drug interactions. Failure to comply with this treatment regimen can lead to XDR-TB, which occurs when patients become resistant to at least one of the injectable second-line anti-TB medicines, amikacin, kanamycin and/or capreomycin, and to a fluoroquinolone. In countries with high proportions of TB drug resistance and those with a large population of TB and HIV/AIDS co-infection, developing rapid detection methods and improving the management of patients with drug-resistant TB is an urgent priority.

Treatment of patients with multidrug-resistant TB is one of the bases of results of drug susceptibility testing (DST) obtained before the treatment initiation. Each dose is given as directly observed therapy (DOT) throughout the treatment. The duration of MDR-TB treatment is 18 to 24 months. The intensive phase being a minimum of six months and the continuation phase 12 to 18 months. Patients with MDR-TB confirmation but not with full DST result are treated with Ethambutol, Pyrazinamid (Z), Kanamycin (Amicacin), Levofloxain (Lfx), Ethionamid (Eto) and Cycloserin (Cs). MDR-TB patients susceptible to both Kanamycin and Quinolons are treated as in the first case. MDR-TB patients susceptible to Kanamycin but resistant to Quinolons are treated with Ethambutol, Pyrazinamid (Z), Kanamycin (Amicacin), Moxifloxacin, Ethionamid, Cycloserin and Para Amino Salicylic Acid (PAS) (WHO/HTM/TB/2008.402) Geneva.

2.3 Prevention and control of multidrug-resistant tuberculosis

Weak national health systems impede basic control and facilitate re-appearance and spread of drug-resistant tuberculosis. Effective control requires appropriate national policies, trained and motivated staff, and quality-assured laboratory- and medicine-supply systems supported by an adequately funded tuberculosis program. All health-care facilities used by patients with symptoms of tuberculosis must be engaged with general and specialized hospitals, academic institutions and the array of diverse private-care providers need to be involved as a priority. A network of patient-friendly health clinics and staff is essential to ensure that treatment is supervised in a supportive manner and is quality-assured, free of cost, and easy to access. If patients discontinue their treatment, there must be mechanisms to trace them and re-establish

treatment. Moreover, informed, motivated and resourced communities can contribute to case finding and adherence support especially in resource-poor settings.

Presently, less than 5% of the estimated cases of multidrug-resistant tuberculosis are being diagnosed (WHO, 2009). Many countries, especially in Africa, lack laboratory capacity to culture Mycobacterium tuberculosis and do drug-susceptibility tests. Laboratory capacity, neglected for a long time, needs rapid expansion under international norms and standards, as part of the strengthening of a broader national public health laboratory system; the Global Laboratory Initiative of WHO and partners is helping to enhance coordination of the response. New technologies that can accelerate the diagnosis of drug-resistant tuberculosis are available, but not yet widely implemented, the main obstacle being lack of an adequate, safe laboratory infrastructure and appropriately trained staff. National programs need policies on where and how to treat drug-resistant tuberculosis cases. In some countries, patients are admitted to hospital for long periods of time, which is labor-intensive and costly, raises important ethical and social issues, and increases the risk of nosocomial transmission if infection control is weak. New models of care enabling safe and effective treatment supplemented by community-based support have proven to be feasible and effective in low-resource settings. To expand treatment services effectively and rapidly, countries will need centers of excellence to ensure adequate capacity building of health-care providers for tuberculosis management. A large proportion of tuberculosis patients are diagnosed and treated in the private sector in many countries and the quality of management is uneven: the patients detected are not notified and their treatment outcomes are unknown. Modes of collaboration with the private sector for care and control including management of multidrug-resistant tuberculosis, in which patients do not have to pay for costs of care, have proved effective in resource-poor settings and are necessary for rapid expansion of multidrug-resistant tuberculosis management. Health ministries should involve the private-care sector in ensuring provision of quality treatment through public–private mix approaches linked with the national tuberculosis program.

Although some individuals who have not had previous TB treatment are infected by MDR-TB, this is not the case for most patients. Many new cases of MDR-TB are created each year by a combination of physician error and poor patient compliance with treatment, which turn fully susceptible organisms, or those with less complex resistance patterns, into MDR-TB. Professor Michael Iseman, the US ‘guru’ of MDR-TB, has shown that two to four errors are needed to turn

a fully susceptible organism into a case of MDR-TB (Iseman MD (1993). Accordingly, there were Ten Commandments for physicians: the first is never to add a single drug to a failing regimen, and the other nine are for the physician to repeat the first commandment nine times to make sure that the message is understood. Support and funding of national TB programs, in which treatment is given as directly observed therapy (DOT), is essential for all persons with TB if at all possible. Physicians should always use evidence-based treatment guidelines and drugs of proven bio-availability. The WHO recommends a 6 month initial treatment regimen of rifampicin, isoniazid, pyrazinamide and ethambutol for 2 months, followed by rifampicin and isoniazid for 4 months (2RHZE–4RH). If the patient fails treatment (positive cultures or sputum smears in months 5 or 6 of treatment) or relapses, an 8 month retreatment regimen is recommended. This consists of streptomycin, rifampicin, isoniazid, pyrazinamide and ethambutol for 2 months, followed by rifampicin, isoniazid, pyrazinamide and ethambutol for 1 month, followed by rifampicin, isoniazid and ethambutol for 5 months (2SRHZE–1RHZE–5HRE), (WHO,1996).

People living with HIV are more susceptible to developing tuberculosis, including drug-resistant tuberculosis. Also, HIV infection greatly increases the fatality rate among multidrug- and extensively drug-resistant tuberculosis. Improved and strengthened collaboration between tuberculosis and HIV programs is required to prevent rapid transmission of drug-resistant tuberculosis and resulting high mortality among communities heavily affected by HIV. To this end, WHO recommended that collaborative tuberculosis/HIV activities should be expanded (Documents WHO/HTM/TB/2004.330). Quality-assured medicines are essential for successful treatment of tuberculosis. Manufacturing processes must meet international standards and the quality of the finished product must be assured. WHO standards for quality medicines are not always observed. Quality-assured fixed-dose combinations, developed as a tool to prevent the emergence of resistance, are not widely used. Inadequate supply of quality-assured second-line medicines has been a major issue. Since 2000, the Green Light Committee, established by WHO and partners, has provided access to medicines that are quality assured to WHO standards, and concessional priced, for projects worldwide that apply WHO guidelines. Concerted action on the part of governments, drug-regulatory authorities, the pharmaceutical industry, and WHO is required to ensure that adequate and uninterrupted supply of quality-assured anti-tuberculosis medicines are available and accessible to all those in need. Combined with a cost per patient

treated that is usually in the range US$ 3000–10 000 the total cost of treating 1.5 million cases amounts to US$ 11 500 million over seven years, rising from US$ 500 million in 2009 to US$ 3100 million in 2015; the latter figure is 43 times the funding available in 2009 and 53% of the total funding required for tuberculosis control. Most funding is required in the European Region (US $7800 million), followed by Asia (US$ 2800 million). In order to mobilize the required funding for improved management of multidrug- and extensively drug-resistant tuberculosis, preparation of country-specific budgets as part of national strategic plans is the first step that needs to be taken. World Health Organization has prepared a planning and budgeting tool for this purpose. Domestic resources need to be accessed especially in middle-income countries. If sufficient domestic funding cannot be mobilized, countries should make full use of resources available from the Global Fund to Fight AIDS, Tuberculosis and Malaria, the International Drug Purchase Facility (UNITAID), and other donor agencies and funding mechanisms.

2.4. Risk factors associated with MDR TB.

2.4.1 Socio-demographic factors

Sex

Doo J. *et al*. (2010) conducted a retrospective study on treatment outcome and mortality among patients with multi-drug resistant tuberculosis in tuberculosis hospitals of the public sectors of Korea. The study aims to evaluate treatment outcomes, mortality and predictors of both in MDR-TB patients at the three TB hospitals in the public sector of Korea. The study covered a total of 202 (135 males and 67 females) MDR TB confirmed patients and obtained that male sex is an independent predictors of poor outcome.

Selamawit H. *et al*. (2013) studied Determinants of multidrug-resistant tuberculosis in patients who underwent first-line treatment in Addis Ababa: a case control study. The study was conducted at St. Peter Hospital and five health centers in Addis Ababa having equal number of cases and controls and aims to assess factors that determine the occurrence MDR-TB among patients who had taken first line anti-TB treatment in Addis Ababa City. The study covered a total of 134 cases and from this 81 (60.5%) of the MDR-TB cases were males and the rest 53 (39.5%) were females. In addition, the study also found that being male was a risk factor for MDR-TB development. A study in Nigeria, Daniel O. *et al*. (2006) Default from tuberculosis treatment programme in Sagamu, Nigeria, also showed that being male was a risk factor for

defaulting from anti-TB medication. Similarly, this study showed that among MDR-TB cases who were defaulters in their first-line TB treatment, 62.5% were males. The association between being male and having MDR-TB could be due to the fact that males have a higher tendency not to adhere to anti-TB treatment than females, thus increasing their risk of developing MDR-TB.

Age

Alena S. *et al*. (2013) conducted a national survey on Multidrug-resistant tuberculosis in Belarus: the size of the problem and associated risk factors. The study covered a total of 1344 cases of MDR TB including 934 new and 410 previously treated cases. The objective of the study was to assess the problem of multidrug-resistant tuberculosis (MDR-TB) throughout Belarus and investigate the associated risk factors by using univariate and multivariate logistic regression analysis. They found that an age of < 35 years was found to be an independent positive risk factor for MDR-TB.

Getachew *et al*. (2013) conducted a retrospective analysis on survival and predictors of mortality among patients under multi-drug resistant tuberculosis treatment in Ethiopia: St. Peter's specialized tuberculosis hospital, Ethiopia. The objective of the study was to assess survival and predictors of mortality among patients under MDR-TB treatment in Ethiopia: St Peter's specialized TB Hospital, Addis Ababa, Ethiopia. The study covered a total of 188 patients and analyzed using Cox proportional hazards model. The study was mainly conducted to assess the socio-demographic and clinical factors for the development of MDR-TB. They found that the majority of MDR-TB patients were younger aged less than 35 years (81.38 %) with median age of 27 years. In addition to that they also revealed that survival of patients under MDR-TB treatment was not associated with age.

Ahmed M. *et al*. (2010) conducted a review on risk factors for multi-drug resistant tuberculosis. The main aim of the review was to determine the risk factors associated with multi-drugs resistant TB (MDR-TB). The review was critically assessing the risk factors that are associated with multi drug resistant tuberculosis. The review suggested that age has been found to be independently associated with drug resistance and there was significantly higher proportion of MDR-TB among the age group of 45-64 years. It also found that MDR-TB was more likely in patients under 65 years, but the association was weak and more heterogeneous in patients under 45years.

Drobniewski, F. *et al*. (1997) conducted a retrospective study on national study of clinical and laboratory factors affecting the survival of patients with multiple drug resistant tuberculosis in the UK. The study covered a total 90 MDR-TB patients for the aims of to describe the clinical, microbiological, molecular epidemiology and treatment of multidrug resistant tuberculosis (MDR-TB) cases in the UK and to determine factors associated with survival. The study was analyzed for three age categories (15-34years, 35-54years and 55^{+} years) resulted that survival was influenced by age. For those aged 15–34 years the median survival could not be calculated because, during the total study period, only 18.4% of the patients in this age group died that is, 81.6% of patients were censored. For cases aged 35–54 years the median survival time was 1379 days (95% CI 649 to 2066) with 50.0% cases censored, and in the 55+ age group the median survival time was 2515 days (95% CI 119 to 2515) with 45.5% censored.

Marital status and Educational level

Socio-Demographic Profile and Determinants of multidrug-resistant tuberculosis in patients who underwent first-line treatment in Addis Ababa: a case control study was conducted by Selamawit H. *et al*. (2012). They used multivariate logistic regression analysis to assess factors that determine the occurrence MDR-TB among patients who had taken first line anti-TB treatment in Addis Ababa City. They found the socio-demographic characteristics of MDR-TB patients and reviled that the marital status and educational level of the patients and found that 63.4% were single, 23.8% were married, 11.9% were Divorced and 0.75% were widow/widower. The educational level of MDR-TB patients were 11.2% were illiterate, 11.9% were read and write, 20.1% were secondary school and 56.7% were educated tertiary and above. Results from the logistic regression analysis revealed that after adjusting for possible confounding factors the study found that MDR-TB development is significantly associated education above 10th grade (AOR = 3.7; 95% CI; 1.1–12.1).

Yanina B. *et al*. (2011) conducted a retrospective national cohort study on Survival of drug resistant tuberculosis patients in Lithuania. The cohort study consisted of a total of 1809 MDR TB confirmed patients. They used Kaplan-Meier survival curves and multivariable Cox regression to describe the epidemiological, clinical and socioeconomic features and survival of a large national cohort of MDR-TB cases and to establish risk factors influencing their survival. The Socio-demographic results indicated that the majority of MDR-TB patients had primary or

secondary education label. The study also showed that adjust only for the effect of any treatment, lower levels of education at the time of MDR-TB diagnosis were independently associated with poorer survival times.

Employment status and Religion

A case-control study was conducted in Botswana, University of Pennsylvania Partnership, Gaborone, by N. M. Zetola *et al*. (2012), to describing the patterns of alcohol use among MDR-TB patients and to determine whether alcohol use is associated with the development of MDR-TB in Botswana. Multivariate logistic regression with MDR-TB case status to clearly identify the impact of alcohol use for the development of MDR-TB and a total of 114 cases were included in the study. The results have shown that most 79 (69.3%) of MDR-TB cases were Employed and 35 (30.7%) of MDR-TB cases were unemployed. Another case-control study was conducted by the tuberculosis lead program medical research council, Pretoria, South Africa together with the International Research and Programs Branch, Division of TB Elimination, Centers for Disease Control and Prevention, Atlanta, Georgia, USA from 1999-2001. The objective of the study was to assess the factors associated with default from multi-drug tuberculosis treatment in South Africa by using multivariate logistic regression analysis the study was conducted to analyze different demographic and clinical variables. The results have shown that most 86 out of 96 (90%) of the cases were Christians and the rest 10 out of 96 (10%) of the cases were non-Christian.

James C.M. Brust *et al*. (2010) employed multivariate logistic regression analysis and sensitivity analysis to study High Treatment Failure and Default Rates for Patients with MDR-TB in KwaZulu-Natal, South Africa. The study was conducted to describe treatment outcomes and determine risk factors associated with unfavorable outcomes among MDR-TB patients admitted to the provincial TB referral hospital in KwaZulu-Natal province, South Africa. The descriptive analysis shows that 277 (22.9%) of MDR-TB patients were Employed, 704 (58.2%) of MDR-TB patients were Unemployed, 228 (18.9%) of MDR-TB patients were Unknown employment status.

2.4.2 Clinical factors

Alcohol use

Zetola N. *et al*. (2012) conducted a case control study on Alcohol use and abuse among patients with multidrug-resistant tuberculosis in Botswana. The study was conducted for the aim of describing the patterns of alcohol use among MDR-TB patients and to determine whether alcohol use is associated with the development of MDR-TB in Botswana. The case-control study considers a total of 114 cases and use conditional multivariate logistic regression with MDR-TB case status to clearly identify the impact of alcohol use for the development of MDR-TB. The results obtained indicated that 40 (35.1%) of MDR-TB patients use alcohol for their life time and the rest 74 (64.9%) were not for their life time. The study also identified alcohol abuse is prevalent among MDR-TB patients in Botswana and could be an important modifiable factor affecting health outcome in this population. In addition they found that MDR-TB subjects have an overall lower lifetime prevalence of any alcohol use than the non-alcohol user. Alcohol abuse is associated with the diagnosis of MDR-TB among people with TB.

Another case-control study was conducted by the tuberculosis lead program medical research council, Pretoria, South Africa together with the International Research and Programs Branch, Division of TB Elimination, Centers for Disease Control and Prevention, Atlanta, Georgia, USA from 1999-2001. The objective of the study was to assess the factors associated with default from multi-drug tuberculosis treatment in South Africa. The study was conducted by using multivariate logistic regression analysis using SAS generalized linear models to assess the association of multiple factors with being a case patient (defaulter). The report showed that 33 (38%) of the cases were use alcohol and 53 (62%) of the cases were not use alcohol. From the alcohol users 3 (3%) were daily drinker and 30 (35%) were occasionally drinker. The study reporting that any alcohol use during treatment was significantly associated with treatment default (UOR 1.7, 95% CI 1.02-2.9) leading that the use of alcohol on an occasional or regular basis was also more commonly reported by cases and 4% of the cases were missed treatment due to alcohol use.

Presence of any chronic disease

Matthew J. *et al*. (2014) conducted a cohort study between January 2009 and December 2011 Diabetes Mellitus, Smoking Status, and Rate of Sputum Culture Conversion in Patients with Multidrug-Resistant Tuberculosis: A Cohort Study from the Country of Georgia. The aim of the study was to determine factors associated with culture conversion rates and estimate the association between DM and the risk of poor treatment outcomes among patients with MDR-TB.

Cox proportional hazards models were used to estimate hazard rate ratios (HR) and 95% confidence intervals (CI) for time to culture conversion. The results obtained show that Among 1,366 patients with sputum culture conversion information, 966 (70.7%) had culture conversion and the median time to conversion was 68 days. The rate of conversion was similar among patients with MDR-TB and DM (adjusted hazard ratio [HR] 0.95, 95% CI 0.71–1.28) compared to patients with MDR-TB only. Another study also conducted by Xiaoliang Yuan *et al*. (2013) Genotyping and clinical characteristics of multidrug and extensively drug-resistant tuberculosis in a tertiary care tuberculosis hospital in Chin. The retrospective study was conducted to analyze the clinical features of patients with MDR and XDR-TB from Jiangxi Province. The result obtained from the study shows that diabetes mellitus was the most common co-morbidity in MDR-TB (16/110, 14.5%) patients.

Kang Y. *et al*. (2013) investigated Impact of diabetes on treatment outcomes and long-term survival in multidrug-resistant tuberculosis. MDR-TB patients newly diagnosed or retreated between 2000 and 2002 and followed for 8-11 years were retrospectively analyzed with respect to the effect of DM as co-morbidity on their treatment outcome and long-term survival. The aim of the study was to assess the impact of DM on treatment outcomes of patients with MDR-TB based on 1,407 patients with MDR-TB, 239 (17.0%) had coexisting DM. The mean age and body mass index were higher in MDR-TB patients with DM [MDR-TB DM positive] than in those without DM [MDR-TB DM negative]. Patients with MDR-TB and a co-morbidity of DM had a significantly lower treatment success rate than those without a history of DM (36.0 vs. 47.2%, p = 0.002). In addition, DM was the negative predictor for MDR-TB treatment success in multivariate analyses [odds ratio 0.51, 95% confidence interval (CI) 0.26-0.99]. Mean survival times were also lower in MDR-TBDM positive than in MDR-TBDM negative patients (102 vs. 114 months, p = 0.001), with DM as a significant predictor of poor long-term survival in multivariate analyses (hazard ratio 1.59, 95% CI 1.01-2.50).

Therapeutic delay and Smear positivity

Dhingra V. *et al*. (2007) conducted on outcome of multi-drug resistant tuberculosis cases treated by individualized regimens at a tertiary level clinic from August 2002 to December 2004. The study was conducted at New Delhi Tuberculosis (NDTB) Centre, a referral center for tuberculosis. All patients with pulmonary MDR-TB attending the outpatient MDR-TB clinic

from August 2002 – December 2004 were included to determine the clinical, radiological and drug resistance profile as well as the factors associated with treatment outcome of Multi-Drug Resistant Tuberculosis (MDR-TB). They used chi-square test to analyze the relation between the outcome of treatment and variables that might influence the outcome and some descriptive parts. The result of the study shows that over 70% patients were already resistant to three or more anti-tuberculosis drugs at presentation. Out of the 27 patients, 13 were cured, 10 defaulted, one died, one is still on treatment with two referred for surgery and 12 patients were smear positive at the time of diagnosis. F Drobniewski *et al*. (2002) also conducted a national study of clinical and laboratory factors affecting the survival of patients with multidrug resistant tuberculosis in the UK and found that 78 out of 90 were sputum smear positive.

Theodros G. *et al*. (2013) conducted survival and predictors of mortality among patients under multi-drug resistant tuberculosis treatment in Ethiopia: st. peter's specialized tuberculosis hospital, Ethiopia. A retrospective analysis of records was conducted from Oct, 2011 - May, 2012 among cohorts of MDR-TB patients in St. Peter's specialized TB hospital that starts treatment from February 2009 and includes 188 MDR TB confirmed patients. The aim of the study was to assess survival and predictors of mortality among patients under MDR-TB treatment in Ethiopia: St Peter's specialized TB Hospital, Addis Ababa, Ethiopia. Relative risks (hazard ratio) with 95% CI and two-sided test of significance was used to measure the association of dependent and independent variables. Survival curves were compared between different exposure groups using log-rank test Survival trend over the follow up time was calculated using the Kaplan Meier (KM) method and the covariates were fitted to Cox proportional hazard regression model and Collett's Model selection approach was used. After fitting the Cox proportional hazard regression model to a set of survival data, the adequacy of the fitted model to the survival data was checked using (Generalized) Cox-Snell residuals and martingale residuals. The result of the socio-demographic and clinical characteristic of MDR-TB cases shows that 93 (49.47%) had therapeutic delay < 1 month and 95 (50.53) had therapeutic delay >= 1 month and showed significant association was therapeutic delay of more than one month. When they compare patient survival who starts treatment a month delay after diagnosed as MDR-TB was observed to have a higher hazard of death (HR= 3.61; 1.41- 9.20, P = 0.007).

HIV status, MDR TB category and number of first line drugs at initiation

Telzak E. *et al*. (1998) dealt with predictors of multidrug-resistant tuberculosis among HIV-infected patients and response to specific drug regimens in Bronx-Lebanon hospital center. The objective of the study was to determine the demographic,behavioral, clinical and geographic risk factor associated with the occurance of MDR-TB among HIV-infected patients and to evaluate the over all survival and clinical response of MDR-TB patients treated with specific drug regimens. X^2 and fisher's exact test (two tail) were employed to determine statistical significance. Logistic regression analysis included all variables that on bivariate analysis were associated with MDR-TB with P value <0.2. for the treatment component,the cumulative probablities of survival (CPS) at 12 and 18 months were calculated by using the actuarial method of life table construction. The result of the study shows that the relationship between culture confirmed MDR-TB and several variables that were hypothesized to be risk factors for MDR-TB. Prior anti-tuberculosis treatment was the only characterstic significantly associated with MDR-TB.Of the 23 patients who had received prior treatment , six (26%) had MDR-TB, compared with 10 (8%)of the 133 patients who had not received prior treatment (odds ratio [OR] 4.3,95% confidence interval [CI] 1.4-13.5, *P*=0.02).on logistic regression analysis, including prior tuberculosis, place of birth and hospitalization within 6 months in a facility in which transmission of MDR-TB was known or suspected to have occurred, only prior tuberculosis treatment was found to be an independent predictor for MDR-TB (OR 4.6, 95% CI 1.4-14.6, *P*=0.01). The only factor associated with the occurance of MDR-TB in their study was previous history of treatment for tuberculosis. Patients with MDR-TB were more than four times more likely to have had prior treatment compared to those with non MDR-TB patients. Previously treated patients had a 30% rate of MDR-TB compared with a 7% rate for those who had not received prior treatment. In a new York citywide survey of tuberculosis isolates from 1991, 19% of patients had isolates that were resistant to at least both isoniazed and rifampin. In that study, a history of tuberculosis treatment was also the strongest predictor of MDR-TB. Previously treated patients had a 30% rate of MDR-TB compared with a 7% rate for those who had not received prior treatment. In a study of 34 HIV-infected patients with MDR-TB reported from Bronx-Lebanone Hospital center, an inner city facility in the south Bronx, the over all response rate was 50%. The median survival for these 34 patients was 315 days. Multivariate analysis revealed that receipt of appropriate therapy, defined as receiving at least two drugs with in vitro activity against the isolate, for at least 2 consecutive weeks, was the only variable

associated with both initial and overall response. For the 20 patients who received two or more active drugs for at least 2 weeks, the CPS at 12 months was 82%.

Songhua C. *et al*. (2013) conducted a case control study on Risk factors for multidrug resistance among previously treated patients with tuberculosis from July through August 2011 in five cities of Zhejiang Province in eastern China. The study aimed to ascertain the risk factors for MDR-TB in this particular population in China and used univariate analysis, the Chi-square test and Fisher's exact test, as well as the Student's t-test (two-sided), was performed to identify risk factors for MDR-TB, and a p-value less than 0.05 was considered statistically significant and a multivariate analysis was performed by forward stepwise (likelihood ratio) multiple logistic regression. In the forward stepwise regression, independent variables with a p-value less than 0.10 were included in the logistic mode. The result of the study show that duration of first treatment of more than 8 months and having had more than three prior episodes of anti-TB treatment were associated with MDR-TB.

Yanina B. *et al*. (2011) conducted a retrospective national cohort study on Survival of drug resistant tuberculosis patients in Lithuania. The national cohort study was conducted to describe the epidemiological, clinical and socioeconomic features and survival of a large national cohort of MDR/XDRTB cases and to establish risk factors influencing their survival. All MDR/XDRTB cases (n=1807, 1736 MDR-TB cases and 71 XDR-TB cases) reported from 2002 to 2008 in Lithuania with a known outcome were included in the survival analysis. They used Kaplan-Meier survival curves and multivariable Cox regression to analyze time until death from any cause during patient's treatment or follow-up, from the time of the first-recorded diagnosis of MDR or XDRTB in the database. They found that a proportion of MDR-TB patients (13%) were resistant to other drugs in addition to resistance to isoniazid and rifampicin. The survival analysis revealed that Median survival for MDR TB patients was 4.0 (95% CI 3.7 to 4.4) years, and for HIV positive versus HIV negative was 1.9 (95% CI 0.4 to 3.5) and 4.9 (95% CI 4.3 to 6.8) years, respectively. They also found that positive or unknown HIV status at the time of MDR-TB diagnosis was independently associated with poorer survival. HIV infection was associated with lower survival; only half survived for up to 1.9 years from MDR-TB diagnosis. They conclude that rapid identification of drug resistance, early administration of appropriate treatment, achievement of high cure rates, adequate infection control measures, expansion of HIV testing

and antiretroviral treatment are necessary to improve patients' survival and prevent further spread of MDR and XDR-TB in Lithuania.

Samuel M. *et al*. (20013) conducted a study on Risk of death among HIV Co-Infected Multidrug Resistant Tuberculosis Patients, Compared to mortality in the general population of South Africa from 200-2004. The study covered a total of 2079 MDR-TB patients to determine excess mortality attributable to HIV among MDR-TB patients in South Africa using relative survival methods. A Poisson-based model was constructed to assess the excess mortality the relative survival time among HIV co-infected MDR-TB patients. Of the 1413 that were tested for HIV, 554 (26.6%) were HIV positive, 859 (41.3%) were HIV negative and the rest 666 (32.1%) were unknown status of HIV. The study also showed that a large number of deaths, 169 (37.98%) deaths, were occurred on HIV positive MDR-TB patients. The relative survival model showed that there was a significant difference in the survival of the patients in terms of their HIV status (p-value < 0.01). For comparative purposes, the results obtained from fitting logistic and proportional hazards regression models on the odds ratio (OR) and hazards ratio (HR) scales are clearly identified. The excess mortality rate was higher in HIV co-infected, than in HIV-negative, patients (adjusted excess hazard ratio, 5.6 [95% CI, 3.2-9.7])

CHAPTER THREE

3. METHODOLOGY

3.1 Introduction

In this study, the methodology to be employed starting with description of study area and population, data and the description of the study variables are discussed. The study concludes discussing different statistical techniques used in the data analysis such as Kaplan-Meier survival analysis, Cox proportional hazard model and parametric survival regression models.

3.2 Study Design

This study was a retrospective cohort study based on data from the MDR-TB ward in Gondar university hospital. The study reviewed patient's MDR-TB charts, intake forms and follow up charts of MDR-TB patients in Gondar University teaching hospital started MDR- TB treatment on August 2010 and followed until to August 2014. Each patient has one medical file containing all MDR-TB notes which includes the patient intake forms and MDR-TB care and follow-up card, prepared by Federal ministry of health (EFMOH) to be uniformly used by clinicians to early identify and document clinical and laboratory variables. Thus, in this study secondary data which was collected from patients follow up records are used. Based on this record of the patients, the variables which were important for the study were selected by using the patient's unique identification number or the laboratory code. This was done by communicating with the nurses and counselors to get the medical record and other important information for the study.

3.3 Study area and its description

The study was conducted at university of Gondar Hospital MDR-TB clinic. The university of Gondar Hospital is one of the oldest academic institutions in Ethiopia. It has been producing a number of health professionals since more than half a century ago. The university is situated at the heart of Gondar city found in Amhara Region, North West of Ethiopia, which located at 727 kilometers away from Addis Ababa. The hospital provides different inpatient and outpatient services to the population in the surrounding area of Gondar town and the nearby words and zones.

3.4 Data

The target population for this study included all patients age ≥18 years with MDR-TB infection at Gondar University hospital MDR-TB clinic (which is located at Gondar town) in Amhara Region and started MDR TB treatment at Gondar hospital from period between August 2010 and August 2014 were followed for a maximum of 24 months. The target population for these studies was patients under MDR-TB treatment at Gondar university hospital started treatment since August 2010. The total number of population started MDR-TB treatment were 224 patients. From the total patients only 146 patients had a complete outcome and the remaining 78 patients were still on treatment and had no outcomes. To conduct the survival time and risk factor analysis, only patients who had known outcomes were taken and no samples were selected due to the fact that increasing the number of participants also increases the precision of the study. So, all patients who had known outcome were included in the study.

3.5 data collection procedures

The secondary data was collected using standardized structured data manipulation form in Gondar university teaching hospital. Relevant data was taken from MDR-TB follow up charts. The data were collected by two professional nurses and one more experienced nurse for supervision was involved. They all were professionals about MDR-TB and no training was given on the definition of variables on the data manipulation form.

3.6 Variables of the study

The response variable

The response or outcome variable in this study is the survival time measured (in months) from the date of the MDR treatment's started until the date of the patient's death or censor (transferred, Dropout, Cured, Lost follow-up, treatment completed and treatment failed).

Predictor Variables

The predictor variables in survival data analysis are called covariates. These are explanatory variables which are assumed to influence the survival time of MDR-TB patients and are given below.

Socio-demographic factors:

- ✓ Sex of the patient
- ✓ Age of the patient
- ✓ Religion

- ✓ Marital status
- ✓ Employment status
- ✓ Educational level

Clinical factors:

- ✓ HIV co-infection
- ✓ MDR category
- ✓ presence of chronic disease
- ✓ clinical complication
- ✓ No. Of resistant drugs at initiation
- ✓ therapeutic delay
- ✓ Alcohol use
- ✓ smoking status
- ✓ smear positivity

Table 3.1: Description of the Independent variables

Variables/factors	Symbol of variable	Categories	
Sex of the patient	X_1	Male	1
		Female	0
Age of the patient	X2	18-34 years	0
		35-54 years	1
		≥55 years	2
Marital status (MRSUS)	X_3	Single	0
		Married	1
		Separated/Divorced	2
		Widow/ Widowed	3
Education level (EDULABL)	X_4	Illiterate	0
		Read and write	1
		Secondary	2
		Tertiary and above	3
Employment status (EMPSUS)	X_5	Employed	0
		Own business	1
		Day laborer	2
		Unemployed	3
Religion	X_6	Orthodox	0
		Muslim	1
		Protestant	2

		Others	3
Therapeutic delay (DELAY)	X7	>= 1 month	0
		< 1 month	1
Number of first line drugs Resistant at initiation(NODRUG)	X8	INH and RIF only	0
		> INH and RIF	1
Smoking status	X9	Yes	0
		No	1
Alcohol use	X10	Yes	0
		No	1
Any clinical complication (Clnccomplicn)	X11	No complication	0
		Pneumonia	1
		Pneumothorax	2
		Hemoptysis	3
		Cor pulmonal	4
MDR category (MDRCAT)	X12	Previously treated	0
		Previously not treated	1
HIV co-infection (HIV)	X13	Positive	0
		Negative	1
		Unknown	2
Presence of any chronic disease (CHRONIC)	X14	No chronic disease	0
		Diabetes mellitus	1
		Myocardial infarction	2
		Asthma	3
		DM and HTN	4
Smear positivity	X15	Positive	0
		Negative	1

3.7 Survival Data Analysis

Survival analysis is the phrase used to describe the analysis of data in the form of time from a well-defined time origin until the occurrence of some particular event or end-point. Survival analysis is used to analyses data in which the time until an event is of interest. The response is often referred to as a failure time, survival time, or event time. Generally, survival analysis is a collection of statistical procedures for data analysis for which the outcome variable of interest is time until an event occurs (time-to-event data), which is always nonnegative and has a positively skewed distribution (Collett, 2003).

One of the most important differences between the outcome variables modeled via linear and logistic regression analyses and the time variable in the survival data is the fact that we may only

observe the survival time partially. The variable time actually records two different things. For those subjects who experienced the event (died), it is the outcome variable of interest, the actual survival time. However, for subjects who have not experienced the even time (cured, lost follow up, transferred, treatment completed and treatment failed). These incomplete observations are referred to as being censored.

There are generally three reasons why censoring might occur: when a subject does not experience the event before the study ends, an individual is lost to follow-up during the study period and a person withdraws from the study for known or unknown reasons. There are three categories of censoring.

i) **Right censoring**: Survival time is said to be right censored when it is recorded from its beginning to a defined time before its end time. This type of censoring is commonly recognized survival analysis and also considered in this study.

ii) **Left censoring**: Survival time is said to be left censored if an individual develops an event of interest prior to the beginning of the study; this is not common in survival studies.

iii) **Interval censoring**: Survival time is said to be interval censored when it is only known that the event of interest occurs within an interval of time but the exact time of its occurrence is not known.

3.7.1 Descriptive Methods for Survival Data

Once we have collected time to event data, descriptive analysis for survival data is our first task to present numerical or graphical (using a survival curve) summaries of the survival times in a particular group. In general, a statistical analysis should begin with a thoughtful and thorough univariate description of the data. Survivor function and hazard function are the two functions of central interest in summarizing survival data.

The Survival Function

Let T be a random variable, which can take any non-negative value, associated with the actual survival times, t (time of death). When the random variable T has a probability distribution with underlying probability density function $f(t)$, the distribution function (cumulative distribution function) of T is then given by:

$F(t)$= P (T<t) =$\int_0^t f(u)du$, t ≥ 0 (3.1)

Which represents the probability that a subject selected at random will have a survival time less than some stated value t. Then, the survival function $S(t)$ is defined as:

$S(t)$= P (T ≥t) = 1 - $F(t)$ (3.2)

The survivor function can therefore be used to represent the probability that an individual survives from the time origin to sometime beyond t. The survival function is the probability that an individual will survive at time t or beyond t, and then relationship between the probability density function f(t) and S(t) will be:

f(t)= $\frac{d(1-S(t))}{dt} = \frac{-\mathrm{dS(t)}}{\mathrm{dt}}$ (3.3)

The Hazard Function

The hazard function is widely used to express the risk or hazard of experiencing the event (death) at some time t, and is obtained from the probability that an individual experiencing the event at time t, conditional on he or she has survived (censoring) to that time. That is, the function represents the instantaneous failure rate for an individual surviving to time t.

The hazard function h(t) is defined by: .

$$\mathrm{h(t)} = \lim_{\Delta \mathrm{t}\to 0} \frac{\mathrm{p\{an\ individual\ fails\ in\ the\ time\ interval(t, t+\Delta t)\backslash it\ survived\ until\ time\ t\}}}{\Delta \mathrm{t}}$$

$$\mathrm{h(t)} = \lim_{\Delta \mathrm{t}\to 0} \frac{\mathrm{P}\,[\mathrm{t} < T < t + \Delta \mathrm{t}\backslash \mathrm{T} \geq \mathrm{t}]}{\Delta \mathrm{t}} \quad (3.4)$$

By applying the theory of conditional probability and the relationship in equation (3.3), the hazard function can be expressed in terms of the underlying probability density function and the survivor function as follows (Collett, 2003).

h(t)= $\frac{f(t)}{S(t)} = -\frac{d}{dt}\{\log S(t)\}$ (3.5)

The corresponding cumulative hazard function H(t)is defined by:

H(t) = $\int_0^t h(u)du = -\log S(t)$ (3.6)

Hence the survival function can be rewritten as

$S(t)$= exp{$-H(t)$}, (3.7)

The hazard rate is not a probability, it is a probability rate. Therefore it is possible that a hazard rate can exceed one in the same fashion as a density function f (t) may exceed one.

Survival data are summarized through estimates of the survival and hazard function. The Kaplan-Meier, Nelson-Aalen and Life Tables are the three commonly used methods for estimating survival and hazard functions (Collett, 2003).

3.7.1.1 Kaplan-Meier Estimator

The first step in the analysis of ungrouped censored survival data is normally to obtain the Kaplan-Meier estimate of the survivor function. The Kaplan-Meier estimator is the standard estimator of the survival function (Collett, 2003). This estimator is also known as the product-limit estimator of the survivor function. The Kaplan-Meier estimator is used to estimate the survival time (time of censoring) of a patient and construct survival curves to compare the survival experience of a patient between different categorical variables.

This method is non-parametric or distribution-free since it does not require specific assumption to be made about the underlying distribution of the survival times.

Suppose the data consist of n survival times $t_1, t_2,\ldots,t_n$ and some of these observations are right-censored times, i.e. for some of the t_j, it is only know that individual j was still censoring at time t_j. Let r be the number of distinct failure times, $r \leq n$, and $t_{(1)} < t_{(2)} < \ldots\ldots < t_{(r)}$ be the ordered failure times. And assume that n_j is the number of patients at censored just before $t_{(j)}$ and d_j is the number of patients who was died at time $t_{(j)}$. Then the Kaplan-Meier estimator of the survival function at time t is given by:

$$\hat{S}(t) = \prod_{j=1}^{k} \left\{ \frac{n_j - d_j}{n_j} \right\} \tag{3.8}$$

for $t_{(k)} \leq t_{(k+1)}$, k=1,2,…,r, with $\hat{S}(t)=1$ for $t < t_{(1)}$.

The standard error of the Kaplan-Meier survival estimator is estimated using Greenwood's formula (Collett, 2003) given as:

$$\text{Se}\{\hat{S}(t)\} = \hat{S}(t)\left\{\sum_{j=1}^{k} \frac{d_j}{n_j(n_j - d_j)}\right\}^{1/2} \tag{3.9}$$

3.7.1.2. Comparing Survival Distributions

After providing a description of the overall survival experience in the study, we turn our attention to a comparison of the survivorship experience in key subjects in the data. When comparing groups of subjects, it is always a good idea to begin with a graphical display of the data in each group. The simplest way of comparing the survival times obtained from two or more groups graphically is to plot the Kaplan-Meier curves for these groups on the same graph. The figure in general shows if the pattern of one survivorship function lies above another, meaning that the group defined by the upper curve lived longer, or had a more favorable survival experience, than the group defined by the lower curve. However, this graph does not allow us to say, with any confidence, whether or not there is a real difference between the groups. The observed difference may be a true difference, but equally, it could also be due merely to chance variation. Assessing whether or not there is a real difference between groups can only be done, with any degree of confidence, by utilizing statistical tests.

A number of statistical tests have been proposed to assess the real difference between groups such as Log-rank, Generalized Wilcoxen, Tarone-Ware test and so on (Hosmer and Lemeshow, 1989). The calculation of each test is based on a contingency table of groups by status at each observed survival time. The general form of these test statistics for the comparison of survival functions between two groups can be defined (Hosmer and Lemeshow, 1989) as follows:

$$Q = \frac{\left[\sum_{j=0}^{r} w_j(d_{1j} - \hat{e}_{1j})\right]^2}{\sum_{j=0}^{r} w_j{}^2 \hat{v}_{1j}} \qquad (3.10)$$

where

r is the number of rank-ordered failure times (event times).

w_j is the weight for censor adjustment at time $t_{(j)}$.

$\hat{e}_{1j} = \frac{n_{1j}\ d_j}{n_j}$ = is the expected number of individuals who experienced an event at time $t_{(j)}$ in group 1, $\hat{v}_{1j} = \frac{n1j\ n2j\ dj(nj-dj)}{nj^2(nj-1)}$ = is the variance of the number of event occurred at time $t_{(j)}$ in group 1, d_{1j} is the observed number of failure (event occur) at time $t_{(j)}$ in group 1, n_{1j} is the number of individuals at risk of event occur in the first group just before time $t_{(j)}$, n_{2j} is the number of individuals at risk in the second group just before time $t_{(j)}$, d_j is the

total number of events occurred at $t_{(j)}$, nj is the total number of individuals at risk before time $t_{(j)}$.

Under the null hypothesis that the two survivorship functions are the same, and assuming that the censoring experience is independent of group, and that the total number of observed events and the sum of the expected number of events is large, *Q* follows a chi-square distribution with one degree of freedom (Hosmer and Lemeshow, 1989).

The most frequently used test is the **log-rank test** (sometimes called the Mantel-Haenszel test and Cox Mantel log-rank test). This test is based on weights equal to one, i.e. $w_j = 1$. The log-rank test is a non-parametric test for comparing two or more independent survival curves. Since it is a non-parametric test, no assumptions about the distributional form of the data need to be made. For the comparison of two groups of survival data the log rank test statistic is given (Hosmer and Lemesho, 1989) by;

$$Q_{LR} = \frac{\left[\sum_{j=0}^{r}\left(d_{1j} - \hat{e}_{1j}\right)\right]^2}{\sum_{j=0}^{r}\hat{v}_{1j}} \qquad (3.11)$$

The statistic Q_{LR} follows a chi-square distribution with one degree of freedom. It tests the extent to which the observed survival times in the two groups deviate from those expected under the null hypothesis of no group differences.

Breslow's test (also known as Gehan's generalised Wilcoxon test) (Collett, 2003) is applicable to data where there is progressive censoring. It is more powerful than the log-rank test when the hazard functions are not parallel and where there is little censoring. It has low power when censoring is high. It gives more weight to early failures.

The Wilcoxon test statistic:

$$Q_{WT} = \frac{\left[\sum_{j=0}^{r} n_j\left(d_{1j} - \hat{e}_{1j}\right)\right]^2}{\sum_{j=0}^{r} n_j^2\hat{v}_{1j}} \qquad (3.12)$$

has a chi-square distribution with one degree of freedom under the null hypothesis that the two survivorship functions are the same, assuming that the censoring experience is independent of the group and the total number of observed events and the sum of the expected number of events

is large . This test uses weights equal to the number of subjects at risk at each survival time, i.e. $w_j=n_j$.

The log-rank test and Wilcoxon test can easily be generalized to the comparison of more than two groups. The statistic for g > 2 groups follows an approximate χ^2 distribution with g−1 degrees of freedom when the null hypothesis of no group difference is true (Collett, 2003).
Of the two, the log-rank test is the more suitable when the alternative to the null hypothesis of no difference between two groups of survival times is that the hazard of death at any given time for an individual in one group is proportional to the hazard at that time for a similar individual in the other group. This is the assumption of proportional hazards, which underlies a number of methods for analyzing survival data. For this reason we use the log-rank test to compare the MDR TB experience among different groups.

3.7.2 Modelling Survival Data

The non- parametric methods (for instance, Kaplan-Meier, Log-rank test) described above can be useful in the analysis of a single sample of survival data, or more groups of survival times. However, in most medical studies that give rise to survival data, supplementary information will also be recorded on each individual. The method (non-parametric method) discussed so far are not suitable for such data set. In order to explore the relationship between the survival experience of individual and explanatory variables, an approach based on statistical modeling can be used (Collett, 2003).

Through a modeling approach to the analysis of survival data, we can explore how the survival experience of a group of individuals depends on the values of one or more explanatory variables, whose values have been recorded for each individual at the time origin.

There are two broad reasons for modeling survival data. One objective of the modeling process is to determine which combination of potential explanatory variables affects the form of the hazard function. Another reason for modeling the hazard function is to obtain an estimate of the hazard function itself for an individual.

A variety of models and methods have been developed for doing this sort of survival analysis using either parametric or semi-parametric approaches. Semi-parametric models are models that parametrically specify the functional relationship between the lifetime of an individual and his

characteristics (demographic, socio-economic, etc.) but leave the actual distribution of lifetimes arbitrary. The most popular semi-parametric model is the proportional hazards model (Collett, 2003).

3.7.2.1 The Cox Proportional Hazards Regression Model

The basic model for survival data is the proportional hazard model. This model was proposed by Cox (1972) and has also come to be known as the Cox regression model. Although the model is based on the assumption of proportional hazards, no particular form of probability distribution is assumed for the survival times. The model is therefore referred to as a semi-parametric model.

The net contribution of socio-demographic variables on experience to an event (death) was assessed by using the Cox proportional hazard model, which combines the features of life table and regression (Cox, 1972).

Cox (1972) proposed a semi-parametric model for the hazard function that allows the addition of covariates, while keeping the baseline hazards unspecified and can take only positive values. With this parameterization, the Cox hazard function is specified as a function of time and the covariates:

$$h(t, \boldsymbol{X}, \boldsymbol{\beta}) = h_0(t).\exp(\boldsymbol{\beta}'\boldsymbol{X}) \tag{3.13}$$

where, $h_0(t)$ is the baseline hazard function that characterizes how the hazard function changes as a function of survival time, $h(t, \boldsymbol{X}, \boldsymbol{\beta})$ represents the hazard function at time t with covariates $\mathbf{X} = (X_1, X_2, \dots, X_p)'$ $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)'$ is a column vector of p regression parameters, $\exp(\boldsymbol{\beta}'\boldsymbol{X})$ characterizes how the hazard function changes as a function of subject covariates. The model (3.13) is referred to as Cox model, or Cox proportional hazards model or simply the proportional hazards model. There are two assumptions of proportional hazards model. The first assumption is the proportional hazard assumption. The assumption of proportional hazards is that the hazard of occurrence of an event at any given time for an individual in one group is proportional to the hazard at that time for an individual in the other group. When there are covariates in the analysis, which are times dependent, this assumption may not hold. This can be verified by considering the hazard ratios of different individuals (Collett, 2003).

For two different individuals with covariates $\mathbf{X_1} = (x_{11}, x_{12,\dots,} x_{1m})'$ and $\mathbf{X_2} = (x_{21}, x_{22,\dots,} x_{2m})'$, the proportion

$$\frac{\mathrm{h}(\mathrm{t},\mathbf{X1},\boldsymbol{\beta})}{\mathrm{h}(\mathrm{t},\mathbf{X2},\boldsymbol{\beta})}=\frac{h_0(t)\exp(\mathbf{X}_1'\boldsymbol{\beta})}{h_0(t)\exp(\mathbf{X}_2'\boldsymbol{\beta})}=\frac{\exp(\mathbf{X}_1'\boldsymbol{\beta})}{\exp(\mathbf{X}_2'\boldsymbol{\beta})}=\exp\big((\mathbf{X}'_1-\mathrm{X}'_2)\boldsymbol{\beta}\big) \tag{3.14}$$

called the hazards ratio, and clearly this ratio is independent of time which means that the log hazard ratio is constant at any given time.

The second assumption is that the relationship between log hazard or log cumulative hazard and a covariate is linear. The Cox proportional hazards model can equally be regarded as linear model, as a linear combination of the covariates for the logarithm transformation of the hazard ratio given by:

$$log\left\{\frac{h(t,\boldsymbol{X},\boldsymbol{\beta})}{h_0(t)}\right\}=log\{e^{\boldsymbol{\beta}'\boldsymbol{X}}\}=\boldsymbol{\beta}'\boldsymbol{X}=\beta_1X_1+\beta_2X_2+\cdots+\beta_pX_p \tag{3.15}$$

The quantity $\boldsymbol{\beta}'\boldsymbol{X}=\beta_1X_1+\beta_2X_2+\cdots+\beta_pX_p$ is called the linear combination of the Cox proportional hazards model.

The hazard function in the Cox model is called semi-parametric function since it does not explicitly describe the baseline hazard function, $h_o(t)$. The survival function of the proportional hazard model is estimated as:

$$S(t,\boldsymbol{X},\boldsymbol{\beta})=e^{-H(t,\boldsymbol{\beta}'\boldsymbol{X})} \tag{3.16}$$

Where, $H(t,X,\beta)$ is the cumulative hazard function at time t for a subject with covariate x.
Since we have assumed that survival time is absolutely continuous, the value of the cumulative hazard function is expressed as:

$$H(t,X,\beta)=H_0(t).\exp(\boldsymbol{\beta}'\boldsymbol{X}) \tag{3.17}$$

Consequently, from the proportional hazards function, we obtained the survivor function given by:

$$S(t,\boldsymbol{X},\boldsymbol{\beta})=[So(t)]^{\exp(\boldsymbol{\beta}'\boldsymbol{X})} \tag{3.18}$$

Where, $H_0(t)$ is the baseline cumulative hazard function and $So(t)$ is the baseline survival function (Collett, 2003).

3.7.2.2 Fitting the Cox Proportional Hazard Regression Model

Fitting the Cox model to observed survival data requires estimating the unknown regression coefficients ($\boldsymbol{\beta}$). Also, the baseline hazard function must be estimated. It turns out that these two components of the model can be estimated separately. The coefficients should be estimated first and the estimates are then used to construct an estimate of the baseline hazard function. The regression coefficients in the proportional hazards Cox model, which are the unknown parameters in the model, can be estimated using the method of maximum likelihood (Collett, 2003).

In Cox proportional hazards model we can estimate the vector of parameters **β** without having any assumptions about the baseline hazard, $h_0(t)$. As a consequence, this model is more flexible and an estimate of the parameters can be obtained easily (Collett, 2003).

Full Maximum Likelihood Estimation

Suppose the survival data based on n independent observations are denoted by the triplet (t_i, δ_i, $\mathbf{X_i}$), i=1, 2... n.

where

t_i is the survival time for the i^{th} individual.

δ_i is an indicator of censoring for the i^{th} individual. given by 0 for censored and 1 for event experience. $\delta = 1$, the actual survival time for the i^{th} individual whose death was occurred , and the exact duration of death was known at the time of the survey) and $\delta = 0$, censored survival time (Censored observations were those in which the patients were not died at the time of survey).

$\mathbf{X_i} = (X_{i1}, X_{i2} \ldots X_{im})'$ is a column vector of m covariates for individual i.

The full likelihood function for right censored data can be constructed as:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} h(ti, \mathbf{X}i, \boldsymbol{\beta})^{\delta i} S(ti, \boldsymbol{X}i, \boldsymbol{\beta}) \qquad (3.19)$$

Where, $h(ti, \mathbf{X}i, \boldsymbol{\beta}) = h_0(ti)\, e^{\boldsymbol{\beta' X}i}$ is the hazard function for the i^{th} individual.

$S(ti, \mathbf{X}i, \boldsymbol{\beta}) = [So(ti)]^{\exp(\boldsymbol{\beta' X}_i)}$ is the survival function for the i^{th} individual. It follows that

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left[h_0(ti)\, e^{\boldsymbol{\beta' X}i}\right]^{\delta i} [So(ti)]^{\exp(\boldsymbol{\beta' X}_i)} \qquad (3.20)$$

The full maximum likelihood estimator of **β** can be obtained by differentiating the right hand side of equation (3.20) with respect to the components of **β** and the base line hazard $h_0(t)$.

This implies that unless we explicitly specify the base line hazard, $h_0(t)$, we cannot obtain the maximum likelihood estimators for the full likelihood. To avoid the specification of the base line hazard, Cox (1972) proposed a partial likelihood approach that treats the baseline hazard as a nuisance parameter and removes it from the estimating equation (Collett, 2003).

Partial Likelihood Estimation

Instead of constructing a full likelihood, we consider the probability that an individual experiences an event at time $t_{(i)}$ given that an event occurred at that time.

Suppose that data are available for n individuals, amongst them there are r distinct failure times and n - r right-censored survival times, and assume that only one individual was died at each ordered failure time, so that there are no ties. The r ordered failure times will be denoted by $t_{(1)} < t_{(2)} < \ldots < t_{(r)}$, so that $t_{(i)}$ is the i^{th} ordered failure time. The set of individuals who are at risk at time $t_{(i)}$ is the i^{th} ordered failure (experiences an event) time, and denoted by R ($t_{(i)}$). And let $\mathbf{X}_{(i)}$ be the vector of explanatory variables for an individual who experiences an event at $t_{(i)}$.

The partial likelihood function is derived by taking the product of the conditional probability of a failure at time $t_{(i)}$, given the number of individuals who are at risk of experiencing the event at time $t_{(i)}$. Then the probability that the j^{th} individual will experience an event at time $t_{(i)}$ is given by: .

$$= \frac{\exp(\boldsymbol{\beta}'\boldsymbol{X}(i))}{\sum_{j\in R\,(t(i))} \exp(\boldsymbol{\beta}'\boldsymbol{X}j)} \qquad (3.21)$$

Where, the summation in the denominator is over all individuals in the risk set.

Thus the partial likelihood is the product over all event time t(i) for i = 1,2,...,r of the conditional probability (3.20) to give the partial likelihood function and can be expressed in the form

$$L_p(\boldsymbol{\beta}) = \prod_{i=1}^{r}\left[\frac{\exp(\boldsymbol{\beta}'\boldsymbol{X}(i))}{\sum_{j\in R\,(t(i))} \exp(\boldsymbol{\beta}'\boldsymbol{X}j)}\right] \qquad (3.22)$$

The product is over the r distinct ordered survival times. The corresponding log-partial likelihood function is given by:

$$log\ L_p(\boldsymbol{\beta}) = \sum_{i=1}^{r}\left\{\boldsymbol{\beta}'\boldsymbol{X}(i) - log\left[\sum_{j\in R\,(t(i))} \exp(\boldsymbol{\beta}'\boldsymbol{X}j)\right]\right\} \qquad (3.23)$$

The partial likelihood derived above is valid when there are no ties in the data set. But in most real situations tied survival times are more likely to occur.

In addition to the possibility of more than one experience an event at a time, there might also be more than one censored observations at a time of event. To handle this real-world fact, partial likelihood algorithms have been adopted to handle ties.

There are three approaches commonly used to estimate regression parameters when there are ties.

These are Breslow (1974), Efron (1977), and Cox (1972) approximations (Collett, 2003).The most popular and easy approach is Breslow's approximation. In many applied settings there will be little or no practical difference among the estimators obtained from the three approximations. Because of this, and since the Breslow approximation is more commonly available, unless stated otherwise, analysis presented in this study will be based on it.

The Breslow Approximation

This approximation is proposed by Breslow and Peto by modifying the partial likelihood which takes the following form

$$L_B(\boldsymbol{\beta}) = \prod_{i=1}^{r} \frac{\exp(\boldsymbol{\beta}' \mathrm{Si})}{[\sum_{j \in \mathrm{R}\,(\mathrm{t(i)})} \exp(\boldsymbol{\beta}\prime \boldsymbol{X} j)]^{d_i}} \qquad (3.24)$$

where s_i is the sum of covariates over d_i subjects at time $t_{(i)}$.

d_i is the number of experienced an event occurred at time $t_{(i)}$.

Now the partial log likelihood of (3.24) is given as .

$$log\ L_B(\boldsymbol{\beta}) = \sum_{i=1}^{r} \left[\boldsymbol{\beta}' \mathrm{Si} - \mathrm{di}\ Log(\sum_{j \in \mathrm{R}\,(\mathrm{t(i)})} \exp(\boldsymbol{\beta}' Xj)) \right] \qquad (3.25)$$

We obtain the Breslow maximum partial likelihood estimator, adjusted for tied observation, by differentiating equation (3.25) with respect to the component of β and setting the derivative equal to zero and solving for the unknown parameters.

When there are no ties, that is, when $d_i = 1$ for each event time, the approximation in equation (3.24) reduce to the likelihood function in equation (3.22).

The maximum likelihood estimates of the regression parameters in the proportional hazards model can be found by maximizing the log-likelihood function in equation (3.23) using

numerical methods. This maximization is accomplished using the Newton-Raphson procedure (Collett, 2003).

The Newton-Raphson procedure is used to maximize the partial likelihood function (3.23) based on the following iterative procedure. An estimate of the vector of **β**-parameters at the $(s+1)^{th}$ cycle of iterative procedure, $\widehat{\boldsymbol{\beta}}_{s+1}$, is given by:

$$\widehat{\boldsymbol{\beta}}_{s+1}=\widehat{\boldsymbol{\beta}}_{s}+\mathbf{I}^{-1}(\widehat{\boldsymbol{\beta}}_{s})\mathbf{U}(\widehat{\boldsymbol{\beta}}_{s}), \qquad \text{for s = 0, 1, 2, ...} \tag{3.26}$$

Where $\mathbf{U}(\widehat{\boldsymbol{\beta}}_{s})$ is the $p \times 1$ vector of first derivatives of the log-likelihood function in equation (3.23) with respect to the **β**-parameters and this quantity known as the vector of efficient scores evaluated at $\widehat{\boldsymbol{\beta}}_{s}$.

$\mathbf{I}(\boldsymbol{\beta}) = -\frac{\partial^2 \log L_p(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_k}$ is the $p \times p$ matrix and known as observed information matrix.

$\mathbf{I}^{-1}(\widehat{\boldsymbol{\beta}}_{s})$ is the inverse of the observed information matrix evaluated at $\widehat{\boldsymbol{\beta}}_{s}$. The variance-covariance matrix of $\widehat{\boldsymbol{\beta}}$, $var(\widehat{\boldsymbol{\beta}})$, can be approximated by the inverse of the information matrix evaluated at $\widehat{\boldsymbol{\beta}}_{s}$ i.e. $\mathbf{I}^{-1}(\widehat{\boldsymbol{\beta}}_{s})$.

The process can be started by taking $\widehat{\boldsymbol{\beta}}_{0}=0$ and continue until the change in the likelihood function is sufficiently small, that is, when $\widehat{\boldsymbol{\beta}}_{s}$ and $\widehat{\boldsymbol{\beta}}_{s+1}$ are sufficiently close together.

3.7.2.3 Variable Selection Procedures

In many settings a variety of explanatory variables are measured and a major question in analyzing such data sets is how to incorporate these variables in the modeling procedure. An initial step in the model selection procedure is to identify a set of explanatory variables that have the potential for being included in the linear component of the proportional hazards model. The methods available to select a subset of the covariates to include in a proportional hazards regression model are essentially the same as those used in the other regression models, like purposeful selection, stepwise (forward selection and backward elimination) and best subsets selection.

When the number of variables is relatively large, it can be computationally expensive to fit all possible models. In this situation, automatic routines for variable selection that are available in many software packages might seem an attractive prospect. These routines are based on forward selection, backward elimination or a combination of the two known as the stepwise procedure.

These automatic routines have a number of disadvantages. Typically, they lead to the identification of one particular subset, rather than a set of equally good ones. The subsets found by these routines often depend on the variable selection process that has been used, that is, whether it is forward selection, backward elimination or the stepwise procedure, and generally tend not to take any account of the hierarchic principle. They also depend on the stopping rule that is used to determine whether a term should be included in or excluded from a model. Thus, instead of using automatic variable selection procedures, the following general strategy for model selection is recommended by Collet (2003).

1. The first step is to fit models that contain each of the variables one at a time. The values of $-2log\hat{L}$ for these models are then compared with that for the null model. The null model is a model in which there are no explanatory variables in the linear component of the hazard model and used to determine which variables on their own significantly reduce the value of -2log$\hat{L}$.
2. The variables that appear to be important from step 1 are then fitted together in a multivariable model. In the presence of certain variables others may cease to be important. Consequently, those variables that do not significantly increase the value of $-2log\hat{L}$ when they are omitted from the model can now be discarded. We therefore compute the change in the value of $-2log\hat{L}$ when each variable on its own is omitted from the set. Only those that lead to a significant increase in the value of $-2log\hat{L}$ are retained in the model. Once a variable has been dropped, the effect of omitting each of the remaining variables in turn should be examined.
3. Variables that were not important on their own, and so were not under consideration in step 2, may become important in the presence of others. These variables are therefore added to the model from step 2, one at a time, and any that reduce $-2log\hat{L}$ significantly are retained in the model. This process may result in terms in the model determined at step 2 ceasing to be significant.
4. A final check is made to ensure that no term in the model can be omitted without significantly increasing the value of$-2log\hat{L}$, and that no term not included significantly reduces $-2log\hat{L}$.

When using this selection procedure, rigid application of a particular significance level should be avoided. In order to guide decisions on whether to include or omit a term, the significance level should not be too small. A level of around 20% - 25% is recommended.

3.7.3 Assessment of Model Adequacy

The adequacy of the model needs to be assessed after the model has been fitted to observed survival data. Model-based inferences depend completely on the fitted statistical model. For these inferences to be valid, the fitted model must provide an adequate summary of the data upon which it is based. Indeed, the use of diagnostic procedures for model checking is an essential part of the modeling process.

As model assumptions checking are based on residuals, we will first introduce the different types of residuals used in survival analysis, and more specifically in the semi-parametric proportional hazards model. Residuals are values that can be calculated for each observation and have the feature that their behavior is known, at least approximately, when the fitted model is satisfactory. Different types of residuals are typical for survival analysis due to the fact that censoring has to be taken into account. Ordinary residuals from linear or generalized linear models are therefore often not applicable.

Cox-Snell Residuals

The residual that is widely used in the analysis of survival data is the Cox-Snell residual, so called because it is a particular example of the general definition of residuals given by Cox and Snell (1968). The Cox-Snell residual for the i^{th} individual, i=1, 2, ..., n, is given by:

$$rc_i = \exp(\hat{\boldsymbol{\beta}}'\mathbf{x}_i)\hat{H}_o(t_i) \qquad (3.27)$$

where $\hat{H}_o(t_i)$ is an estimate of the baseline cumulative hazard function at time t_i, the observed survival time of that individual. Note that from equation (3.27), the Cox-Snell residual, rc_i, is the value of $\hat{H}_i(t_i) = -log\hat{S}_i(t_i)$, where $\hat{H}_i(t_i)$ and $\hat{S}_i(t_i)$ are the estimated values of the cumulative hazard and survivor functions of the i^{th} individual at t_i.

In the argument, if the model fitted to the observed data is satisfactory, then the model-based estimate of the survivor function for the i^{th} individual at t_i, the survival time of that individual, will be close to the corresponding true values $S_i(t_i)$. If the observed survival time for an individual is right-censored, the corresponding value of the residual is also right censored. The

residual will therefore be a censored sample from the unit exponential distribution, and a test of this assumption provides a test of model adequacy. Furthermore, since the Cox-Snell residuals are assumed to have an exponential distribution when an appropriate model has been fitted, they have a highly skewed distribution.

Censored observations lead to residuals that cannot be regarded on the same footing as residuals derived from uncensored observations. Therefore to take an account for censoring, the modified Cox-Snell residual, known as martingale residual, is used.

Martingale Residuals

For the i^{th} individual, i=1,2,…,n, the martingale residuals are given by:

$$rm_i = \delta_i - rc_i \tag{3.28}$$

Where, δ_i is event indicator, which takes the value zero if the observed survival time of the i^{th} individual is censored and unity if it is uncensored.

Martingale Residuals take values in (-∞; 1] and are always negative for censored observations. In large samples, the martingale residuals are uncorrelated with one another and have expected value of zero. However, the martingale residuals are not symmetrically distributed about zero.

Schoenfeld Residuals

Two disadvantages of Cox–Snell residuals and Martingale residuals are that they depend heavily on the observed survival time and require an estimate of the cumulative hazard function. Both of these disadvantages are overcome in a residual proposed by Schoenfeld (1982). These residuals were originally termed partial residuals, but are now commonly known as Schoenfeld residuals.

Schoenfeld residual differs from those considered previously in one other important respect. This is that there is not a single value of the residual for each individual, but a set of values, one for each explanatory variable included in the fitted Cox regression model.

The i^{th} partial or Schoenfeld residual for X_j, the j^{th} explanatory variable in the model, is given by:

$$rp_{ji} = \delta_i\{x_{ji} - \hat{\alpha}_{ji}\}, \tag{3.29}$$

Where, X_{ji} is the value of the j^{th} explanatory variable, j=1, 2,…,p, for the i^{th} individual in the study, and if individuals in the risk set are indexed by l, then:

$$\hat{\alpha}_{ji} = \frac{\sum_{l \in R(t_i)} X_{ji} \exp(\hat{\boldsymbol{\beta}} \boldsymbol{X}_1)}{\sum_{l \epsilon R(t_i)} \exp(\hat{\boldsymbol{\beta}} \boldsymbol{X}_1)} \tag{3.30}$$

And $R(t_i)$ is the set of all individuals at risk at time of t_i.

Schoenfeld residuals are also used to check the proportionality of the covariates over time that is to check the validity of the proportional hazards assumption. If the model fits well then the residuals are randomly distributed without any systematic pattern around the zero line, reference line.

3.7.3.1 Overall Goodness of Fit

One method of checking goodness of fit of the model is to use R^2. In proportional hazards regression model as in all regression analyses there is no single, simple method of calculating and interpreting R^2, because in Cox proportional hazards model, R^2 depends on the proportion of the censored observations in the data. A perfectly adequate model may have what, at face value, seems like a terribly low R^2 due to high percent of censored data (Hosmer and Lemeshow, 1998). The measure of goodness of fit R^2_p based on partial likelihood is given by:-

$$R_p^2 = 1 - \exp\left[\frac{2}{n}(L_O - L_p)\right] \quad (3.31)$$

where,

L_o is the log partial likelihood for empty/null model, the model with no covariates.

L_p is log of partial likelihood for the fitted model with p covariates, and n is the total number of observations in the model.

Under the assumption of proportional hazards, there are three different tests for model assessment (the significance of the coefficients): the partial likelihood ratio test, the Wald test and the score test (Hosmer and Lemeshow, 1998).

Partial Likelihood Ratio Test

Partial likelihood ratio test is the easiest test to compute and the best of the three tests for assessing the significance of the fitted model (for testing the significance of a subset of q explanatory variables from p explanatory variables) (Hosmer and Lemeshow, 1998).

The partial likelihood ratio test statistic, G_{LR}, is given by:

$$G_{LR} = 2\{logL_p(\widehat{\boldsymbol{\beta}}) - logL_p(\boldsymbol{\beta}_0)\} \quad (3.32)$$

Where, $logL_p(\widehat{\boldsymbol{\beta}})$ is the log-partial likelihood evaluated at $\widehat{\boldsymbol{\beta}}$ and $logL_p(\boldsymbol{\beta}_0) = logL_p(\mathbf{0})$ is the log-partial likelihood evaluated at $\boldsymbol{\beta}_0 = \mathbf{0}$.

Under the null hypothesis, H_o: $\boldsymbol{\beta}_{q\times 1} = \boldsymbol{\beta}_0 = \mathbf{0}_{q\times 1}$, that all q coefficients are simultaneously equal to zero, and under mathematical regularities and large sample size conditions G_{LR} follows a chi-square distribution with q degree of freedom, X^2_q.

Wald test

The Wald test is used to check the overall goodness of fit as well as checking the significance of each parameter of the model.

Under the hypothesis, H_o: $\boldsymbol{\beta}_0 = \mathbf{0}_{q\times 1} = (0, 0, 0, \ldots, 0)'$ vs H_1: at least one $\beta_i \neq 0$, $\widehat{\boldsymbol{\beta}}$ will be asymptotically normally distributed with mean $\mathbf{0}$ covariance matrix estimated by $\widehat{\mathrm{V}}\mathrm{ar}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{I}(\widehat{\boldsymbol{\beta}})^{-1}$. Then, the Wald test statistic, G_w , given by:

$$G_W = (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)' \boldsymbol{I}(\widehat{\boldsymbol{\beta}})^{-1} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \widehat{\boldsymbol{\beta}}' \boldsymbol{I}(\widehat{\boldsymbol{\beta}})^{-1} \widehat{\boldsymbol{\beta}} \tag{3.33}$$

follows a chi-square distribution with q degree of freedom, $X^2(q)$.

Score test

The score test statistic, to test H_0: $\boldsymbol{\beta}_0 = \mathbf{0} = (0, 0, 0\ldots.0)'_{\,q\times 1}$ is defined as

$$G_S = U(\boldsymbol{\beta}_0)' \boldsymbol{I}(\boldsymbol{\beta}_0)^{-1} U(\boldsymbol{\beta}_0) \tag{3.34}$$

Where, $U(\boldsymbol{\beta}_0)$and $\boldsymbol{I}(\boldsymbol{\beta}_0)^{-1}$are the score vector and inverse of the observed information matrix evaluated at $\boldsymbol{\beta}_0$. Under null hypothesis and for large sample G_S is asymptotically distributed as chi-squared with q degree of freedom, X^2_q.

3.7.3.2 Checking for the Proportional Hazards Assumption

Once a suitable set of covariates has been identified, In order to use the Cox model, it is wise to check each covariate to ensure that the proportional hazards assumption is valid. This is a critical assumption of proportional hazards model and must be checked for each covariate. To assess the proportional hazards assumption we examine the extent to which the estimated hazard curves for each level of strata of a covariate are equidistant over time. Different tests and graphical techniques have been developed to check whether the proportional hazards assumption holds.

The Grambsch-Therneau test of non-proportionality uses partial residuals for the test of proportional hazards assumption. In order to use this test for the i^{th} covariate, $\mathbf{X}_i$, Grambsch and Therneau (1994) propose a time-varying coefficient as:

$$\beta_i(t) = \beta_i + \gamma_i g_i(t) \quad (3.35)$$

Where, $\beta_i(t)$ is time varying coefficient, β_i is constant, $g_i(t)$is some specified function of time, usually $g_i(t) = \ln(t)$. The Cox proportional hazard model for time varying coefficient with $g_i(t) = \ln(t)$ becomes

$$h\big(t, \mathbf{X}_i, \beta_i(t)\big) = h_0(t)\exp(\beta_i(t)\mathbf{X}_i)$$

Substitute $\beta_i(t) = \beta_i + \gamma_i g_i(t)$ gives

$$\begin{aligned} h\big(t, \mathbf{X}_i, \beta_i(t)\big) &= h_0(t)\exp(\beta_i + \gamma_i \ln(t))\mathbf{X}_i) \\ &= h_0(t)\exp(\beta_i \mathbf{X}_i + \gamma_i \ln(t)\, \mathbf{X}_i \end{aligned} \quad (3.36)$$

This looks like the proportional hazards model where the interaction term, $\mathbf{X}_i \ln(t)$ is included in the model in addition to the main effect $\mathbf{X}_i$. To test the significance of the interaction term $\mathbf{X}_i \ln(t)$, that is, $H_0 : \gamma_i = 0$ against $H_1 : \gamma_i \neq 0$ we can use Wald and/or Likelihood Ratio tests. If $\gamma_i = 0$ is not rejected, $\beta_i' s$ are not time varying coefficients and hence the proportional hazards assumption is satisfied. If $\gamma_i = 0$ is rejected then the proportional hazards assumption is not satisfied.

The Schoenfeld residuals graphical technique can be used to assess Cox model assumptions. For greater diagnostic power the scaled schoenfeld residual is preferred. The scaling can be done on the variance of the i^{th} subject Schoenfeld residuals. If the plot of scaled Schoenfeld residuals versus the logarithm of time is a random, smooth, straight line about zero the proportional hazards assumption will be satisfied.

3.7.3.3 Checking for Linearity of Continuous Covariates

The assumption of linearity can be checked by using the plot of martingale residuals. The plot of martingale residuals obtained from fitting the model, excluding the covariate whose functional form needs to be determined, against the excluded covariate display the functional form required for the covariate. If the resulting plot is random showing no systematic pattern and the smoothed plot is a horizontal straight line. This indicates that the covariate is linear in the model.

3.7.3.4 Extensions of the Proportional Hazards Model

We have used a proportional hazards model with a common unspecified baseline hazard function where all the study covariates had values that remained fixed over the follow-up period (i.e. the effect of a given covariate not changing over time). If the assumption of proportionality is

violated, the simple Cox regression model is invalid and more complicated analyses such as the stratified Cox regression model or the extended Cox regression model is required. Then to accommodate non-proportionality assumption one can apply stratified proportional hazards model in which the stratification in most cases is done by using a covariate fixed by design (Nihal and Tekin, 2007).

The stratified Cox regression model is a modification of the Cox regression model by the stratification of a covariate that does not satisfy the proportional hazards assumption. Covariates that are assumed to satisfy the proportional hazards assumption are included in the model, whereas the predictor being stratified is not included (Nihal and Tekin, 2007).

Let k covariates fail to satisfy the proportional hazards assumption, and p covariates satisfy proportional hazards assumption. The covariates not satisfying the proportional hazards assumption are denoted by Z_1, Z_2, … ,Z_k and the covariates satisfying the proportional hazards assumption are denoted by X_1, X_2, …, X_p. To form the stratified Cox regression model, a new variable is defined from z variables and denoted by z*. The stratification variable z* has k* categories, where k* is the total number of combinations (strata) formed after categorizing each of z's. The stratified cox regression model is defined (Klein et al, 1997) as:

$$h_g(t, \boldsymbol{X}, \boldsymbol{\beta}) = h_{0g}(t)\exp(\boldsymbol{\beta}'\boldsymbol{X}) \tag{3.37}$$

Where, the subscript g represents the strata. The strata are the different categorizations of the stratum variable.

The variable z* is not implicitly included in the model, whereas the **x**'s which are assumed to satisfy the proportional hazards assumption are included in the model. The baseline hazard function, $h_{0g}(t)$, is different for each stratum. However, the coefficient vector $\boldsymbol{\beta}$ is the same for each stratum. Since the coefficients of the x's are the same for each stratum, the hazard ratios are same for each stratum.

The form of the partial likelihood for the g^{th} stratum is identical to the partial likelihood used in proportional hazards model, but it includes an additional subscript, g, indicating the stratum. The contribution to the partial likelihood for the g^{th} stratum is

$$L_{gp}(\boldsymbol{\beta}) = \prod_{i=1}^{n_g}\left[\frac{\exp(\beta' X_{gi})}{\sum_{j\epsilon R(t_{gi})}\exp(\beta' X_{gi})}\right]^{\delta_{gi}} \tag{3.38}$$

Where, n_g: the number of observations in the g[th] stratum

t_{gi}: The i[th] observed value of time in g[th] stratum

$\delta_{gi:}$ The value of the censoring indicator associated with t_{gi}

$R(t_{gi})$: The risk set for subjects in stratum *s* at time t_{gi}

X_{gi} : The vector of p -covariates for subject i in stratum g .

The full stratified partial likelihood is obtained by multiplying the contributions to the likelihood, namely

$$L_p(\boldsymbol{\beta}) = \prod_{g=1}^{g} L_{gp}(\boldsymbol{\beta}) \qquad (3.39)$$

The maximum stratified partial likelihood estimator of the parameter vector, **β**, is obtained by solving the p equations obtained by differentiating the log ($L_{gp}(\boldsymbol{\beta})$) with respect to the p unknown parameters and setting the derivatives equal to zero. Finally model building and model assessment is the same as that of proportional hazards model.

3.7.3.5 Interpretation of the Coefficients of the Final Cox-Regression Model

When the proportional hazards Cox model is used in the analysis of survival data, the coefficient of the explanatory variables in the model can be interpreted as the logarithm of the ratio of the hazard of death of the i^{th} group to the baseline (reference group) hazard. The higher the hazard ratio the lower is the survival probability, and vice versa. If for an exposed group the hazard ratio is high, the survival probability would be equivalently low. In addition, the parameter $\beta'_j s$ can be interpreted as the rate of change in log (HR) per unit change in the j^{th} covariate.

3.8 Parametric Survival Regression Models

In previous topics it was focused entirely on the use of semi-parametric model, proportional hazards Cox regression model, in the analysis and prediction of the survival time of MDR TB patients. The basis of this method is to avoid having to specify the hazard function completely. However, there may be setting in which the distribution of the survival time is in specific parametric distribution that justifies the use of a fully parametric model to better address the goal of the analysis. A parametric survival model assumes that the survival time follows a known distribution. The popularity of this approach is due to the fact that plausible models may be

easily fit, evaluated and interpreted. Many models using different distributions have been developed. Some of most common survival models are:

3.8.1. The Exponential Survival Regression Model

The simplest model for the hazard function is to assume that it is constant over time. The hazard of death at any time after the time origin the study is then the same, irrespective of the time elapsed (Collett, 2003). Under this model, the hazard function is written as:

$$h_0(t) = \lambda \tag{3.40}$$

From the constant baseline hazard function, the corresponding survivor function is:

$$S_0(t) = \left\{-\int_0^t \lambda du\right\},$$

$$= exp(-\lambda t) \tag{3.41}$$

And so the implied probability density function of the survival times is

$$f_0(t) = \lambda\, exp(-\lambda t) \tag{3.42}$$

This is the probability density function of a random variable T that has an exponential distribution with a mean of λ^{-1}.The parameter λ with $\lambda > 0$, is often called the intensity. The median event time can be obtained by solving the equation $S_0(t_{0.5}) = 0.5$ which leads to $t_{0.5} = log\ 2/\lambda$. More generally, the p^{th} quantile can be obtained by solving the equation $S(t_p) = 1 - p$ and thus $t_p = \frac{-\log(1-p)}{\lambda}$ (3.43)

The main feature of the exponential distribution is thus that the instantaneous hazard does not vary over time. Another important property is the lack of memory property. Consider a random variable T ~ Exp (λ). We now study the survival function of a subject conditional on having survived up to time t_0, the excess survival time is described by the same exponential distribution with constant hazard rate λ. An empirical check for this distribution for a set of survival data is provided by plotting the log of the survival function estimate versus t. Such a plot should resemble a straight line through the origin, as $log\ S_0\ (t) = -\lambda t$ if the exponential distribution assumption holds.

3.8.1.1 Fitting the exponential survival regression model

In the parametric setting, estimates of the parameters are obtained by maximizing the likelihood function. The survival likelihood for survival data with event times and right censored data is generally given by:

$$L = \prod_{i=1}^{n} (f(xi))^{\delta i} \, (S(xi))^{1-\delta i} \tag{3.44}$$

which leads for exponentially distributed event times to:

$$L = \prod_{i=1}^{n} (\lambda \, exp(-\lambda xi))^{\delta i} \, (exp(-\lambda xi))^{1-\delta i}$$

$$= \prod_{i=1}^{n} \lambda^{\delta i} \exp(\lambda xi) \tag{3.45}$$

By differentiating the log likelihood function with respect to λ and equating this expression to zero leads to the maximum likelihood estimator

$$\lambda = \frac{d}{\sum_{i=1}^{n} xi} \tag{3.46}$$

3.8.2. The Weibull Survival Regression Model

The Weibull distribution is a generalization of the exponential distribution. However, unlike the exponential distribution, it does not assume a constant hazard rate and therefore has broader application. The distribution was proposed by Weibull (1939) and its applicability to various failure situations discussed again by Weibull (1951). The baseline hazard function for Weibull distributed event times is given by:

$$\mathrm{h}_0(\mathrm{t}) = \lambda \rho t^{\rho-1} \tag{3.47}$$

It follows that the survival function for the Weibull distribution is given by:

$$S_0(t) = \exp(-\lambda t^{\rho}) \tag{3.48}$$

and the density function is

$$f_0(t) = \lambda \rho t^{\rho-1} \exp(-\lambda t^{\rho}) \tag{3.49}$$

with λ, $\lambda > 0$, the scale parameter and ρ, $\rho > 0$, the shape parameter.

The median event time can be obtained by solving the equation $S(t_{0.5})$ = 0.5 which leads to $t_{0.5} = \left(\frac{log2}{\lambda}\right)^{1/\rho}$. More the p^{th} quantile can be obtained by solving the equation $S(t_p) = 1 - p$ and thus $t_p = \left(\frac{-\log(1-p)}{\lambda}\right)^{1/\rho}$ (3.50)

The shape of the hazard function critically depends up on the values of ρ.

If ρ < 1: hazard decreases monotonically with time

If ρ < 1: hazard increases monotonically with time

If ρ = 1: constant hazard (equivalent to exponential distribution)

The Weibull hazard model can be generally presented as

$$h_i(t) = \mathrm{h}_0(\mathrm{t}) \exp(\beta' x_i) \quad (3.51)$$

$$S_i(t) = (\exp -\lambda \exp(\beta' x_i) t^{\rho}) \quad (3.52)$$

$$f_i(t) = \lambda\rho t^{\rho-1} \exp(\beta' x_i)(\exp -\lambda \exp(\beta' x_i) t^{\rho}) \quad (3.53)$$

with $\mathrm{h}_0(\mathrm{t}) = \lambda\rho t^{\rho-1}$ and β a p × 1 vector containing the parameters. The event time of the i^{th} subject is then characterized by the Weibull distribution with scale parameter $\lambda \exp(\beta' x_i)$ and shape parameter ρ. Thus, all subjects share the shape parameter but differ with respect to their scale parameter. The model assumes that individual i and j with covariates X_i and X_j have proportional hazard function of the form:

$$\frac{h(t;X_i)}{h(t;X_j)} = \frac{exp(\beta' X_i)}{exp(\beta' X_j)} = exp\left(\beta'(X_i - X_j)\right) \quad (3.53)$$

The quantities $\exp(\beta)$ can be interpreted as hazard ratios.

3.8.2.1 Fitting the Weibull Survival Regression model

The survival likelihood for Weibull distributed survival data with event times and right censored data is generally given by

$$L = \prod_{i=1}^{n} \left\{ \left(t_i \lambda\rho {x_i}^{\rho-1} \exp(-\lambda {x_i}^{\rho})\right)^{\delta_i} \left(\exp(-\lambda {x_i}^{\rho})\right)^{1-\delta_i} \right\} \quad (3.54)$$

resulting in the log likelihood function

$$l = d + \log(\lambda\rho) + (\rho - 1)\sum_{i=1}^{n} \delta_i log x_i - \lambda \sum_{i=1}^{n} x_i^{\rho} \quad (3.55)$$

with d the total number of events. Maximum likelihood estimators can be obtained by equating the first derivatives of l with respect to λ and ρ to zero and we get.

$$\hat{\lambda} = \frac{d}{\sum_{i=1}^{n} x_i^{\hat{\rho}}} \text{ and}$$

$$\frac{d}{\hat{\rho}} + \sum_{i=1}^{n} \delta_i log x_i - \frac{d}{\sum_{i=1}^{n} x_i^{\hat{\rho}}} \sum_{i=1}^{n} x_i^{\rho} log x_i = 0 \quad (3.56)$$

which is nonlinear in $\hat{\rho}$ and can only be solved by a numerical procedure such as the Newton Raphson algorithm.

3.8.3. Model Selection in Parametric Survival Regression Models

To select the model that can predict the survival of MDR TB patient, we have two methods. The first is graphical approach. For this method the cox-Snell plot is the common one. It is a graph of the minus ln of Kaplan-Meier plotted against the cox-Snell residual values. It is used to determine how well a specific distribution fits to the observed data. This plot would be approximately linear if the specified theoretical distribution is the correct model. Easy fit displays the reference diagonal line along which the graph points should fall along with the goodness of fit tests; the distribution plots can be helpful to determine the best fitting model. The fundamental difference of this approach is that it is quite subjective to come on conclusion while the goodness of fit tests are "exact" in the sense that the results do not depend on the researcher (provided that the tests are performed correctly), using plot is a more empirical way to use in model selection. Akaikie (1974) proposed an informative criterion (AIC) statistic to compare different models and/or models with different numbers. For each model the value is computed as:

$$AIC = -2\log likelihood + 2(p+1+k) \quad (3.52)$$

Where, p denotes the number of covariates in the model without including the constant term and k is the number of parameters minus one *i.e.* s=0 for the Exponential regression and k=1 for Weibull regression models. According to the criterion, a model with small AIC value will be considered as it fits for the data.

CHAPTER FOUR

4. RESULT AND DISCUSSION

4.1 Descriptive survival analysis

4.1.1 Socio-Demographic Characteristics

The study included 146 MDR TB patients, who had known outcome, with 42 (28.8%) of the patients were died and the rest 104 (71.2%) of the patients were censored at the time of the study. Table 4.1 shows the total of 146 MDR TB patients, 60.3% were males and 39.7% were females. Regarding MDR TB patients' age at the time of diagnosis, 36.9% patients were between 18-34 years old, 37.7 % patients were between 35-54 years old and the remaining 25.4% patients were more than 54 years old. When we come to marital status of the MDR TB patient 36.9%, 46.6%, 10.3% and 6.2% of the patients were Single, Married, Separated/ divorced and Widow/Widowed respectively. With regard to educational attainment, about 22.6% of the patients were Illiterate, 46.6% had only read and write, 28.1% of the MDR TB patients were Secondary and the remaining 13.0% were Tertiary and above. There were 13.0% of employed MDR TB patients, 28.8% were has their own business, 8.9% were Day laborer and most 49.3% of MDR TB patients were unemployed. 92.5% of MDR TB patients were orthodox and the proportion of protestant and Muslim respectively was 1.3% and 6.2%.

Table 4.1: Summary of some important socio-demographic characteristics of MDR TB patients at Gondar University hospital

Covariates	Category	Censored (%)	Event/ Death (%)	Total (%)
Sex of the patient	Male	63 (71.6)	25 (28.4)	88(60.3)
	Female	41 (70.7)	17 (29.3)	58(39.7)
Age of the patient	18-34 years	49 (90.7)	5 (9.3)	54 (36.9)
	35- 54 Years	33 (60)	22 (40)	55 (37.7)
	>=55 years	22 (59.5)	15 (40.5)	37 (25.4)
Marital status of the patient	Single	45 (83.3)	9 (16.7)	54 (36.9)
	Married	43 (63.2)	25 (36.8)	68 (46.6)
	Separated/ Divorced	10 (66.7)	5 (33.3)	15 (10.3)
	Widow/Widowed	6 (66.7)	3 (33.3)	9 (6.2)

Educational label of the patient	Illiterate Read and Write Secondary Tertiary and above	24 (72.7) 35 (66) 32 (78) 13 (68.4)	9 (27.3) 18 (34) 9 (22) 6 (31.6)	33 (22.6) 53(36.3) 41 (28.1) 19 (13.0)
Employment status of the patient	Employed Own Business Day laborer Unemployed	14 (73.7) 27 (64.3) 9 (69.2) 54 (75)	5 (26.3) 15 (35.7) 4 (30.8) 18 (25)	19 (13.0) 42 (28.8) 13 (8.9) 72 (49.3)
Religion of the patient	Orthodox Muslim Protestant	96 (71.1) 6 (66.7) 2 (100)	39 (28.9) 3 (33.3) 0 (0)	135 (92.5) 9 (6.2) 2 (1.3)

4.1.2 Clinical characteristics

The results displayed in Table 4.2 regarding to therapeutic delay 37.7 % of MDR TB patients have started treatment after one month of diagnosis and the remaining 62.3% have started before one month of diagnosis. 47.3 of patients were take only INH and RIF at initiation and most 52.7% of the patients were taken more than INH and RIF at initiation. With regard to addiction, 27.4 % were smokers and 72.6% of the patients were nonsmokers. Most 63.7 % of the patients were alcohol users and 32.3 % of the patients were not alcohol users. Majority 65.8 % of MDR TB patients had no clinical complication and the rest 34.2% were with different clinical complication, 8.9 %, 11.7 %, 6.8 % and 6.8 % were with Pneumonia, Pneumothorax, Hemoptysis and Cor pulmonal clinical complications respectively. Among the MDR TB patients, 80.1% were previously treated and 19.9 % were previously not treated and 13.0 % of MDR TB patients were HIV positive and most 87.0 % of MDR TB patients were HIV negative. With regard to presence of any chronic disease as co-infection, Majority 77.4 % of MDR TB patients had no any chronic disease and the rest 22.6 % were with different chronic diseases, 12.3 %, 4.1 % and 6.2 % were with Diabetes mellitus, Myocardial infarction and Asthma co-infections respectively. Majority 76.7% of the patients were smearing positive and the remaining 23.3 % of the patients were smear negative at the time of diagnosis.

Table 4.2: Summary of some important clinical characteristics of MDR TB patients at Gondar University hospital

Covariates	Category	Censored (%)	Event/ Death (%)	Total
Therapeutic Delay	>= 1 month < 1 month	30 (54.5) 74 (81.3)	25 (45.5) 15 (18.7)	55 (37.7) 91 (62.3)
Number of first line drugs	INH and RIF only More than INH and RIF	37 (53.6) 67 (87)	32 (46.4) 10 (13)	69 (47.3) 77 (52.7)
Smoking status	Yes No	12 (30) 92 (86.8)	28 (70) 14 (13.2)	40 (27.4) 106 (72.6)
Alcohol use	Yes No	54 (58.1) 50 (94.3)	39 (41.9) 3 (5.7)	93 (63.7) 53 (32.3)
Any clinical complication	No complication Pneumonia Pneumothorax Hemoptysis Cor pulmonal	81 (84.4) 4 (30.8) 10 (58.8) 5 (50) 4 (40)	15 (15.6) 9 (69.2) 7 (41.2) 5 (50) 6 (60	96 (65.8) 13 (8.9) 17 (11.6) 10 (6.8) 10 (6.8)
MDR Category	Previously Treated Previously not Treated	97 (82.9) 7 (24.1)	20 (17.1) 22 (75.9)	117 (80.1) 29 (19.9)
HIV Co-infection	Positive Negative	3 (15.8) 101 (79.5)	16 (74.2) 26 (20.5)	19 (13.0) 127 (87.0)
Presence of any chronic disease	No chronic disease Diabetes mellitus Myocardial infarction Asthma	93 (82.3) 10 (55.6) 0 (0) 1 (11.1)	20 (17.7) 8 (44.4) 6 (100) 8 (88.9)	113 (77.4) 18 (12.3) 6 (4.1) 9 (6.2)
Smear positivity	Positive Negative	89 (73.6) 10 (60	32 (26.4) 15 (40)	112 (76.7) 25 (23.3)

4.2 Comparison of Kaplan Meier Survival Curves

The survival status of the total of study subjects 42 (28.7%) were died and 104 (71.3%) were censored at August; 2014. The minimum duration of follow up was 1 Month and maximum was 37 months. Table 1 of **Appendix B** showed that the probability of time to death was high in the first months, which relatively decreases as follow up time increases. During the first month of MDR TB treatment, the maximum (97.95%) probability of death was observed with a standard error of 0.0117, at the 6^{th} months the probability of continuing death of MDR TB patient treated with MDR TB treatment was 88.21% with a standard error of 0.0269, at the 12^{th} months the probability of continuing death of MDR TB patient treated with MDR TB treatment was

80.35% with a standard error of 0.0333 and at the 24th months the probability of continuing death of MDR TB patient treated with MDR TB treatment was 69.68% with a standard error of 0.0394 for the follow-up period of time. In addition, the plot of overall Kaplan-Meier estimate indicate that for MDR TB patients who were treated at University of Gondar Hospital, death relatively decreases as follow up time increases (see figure 4.1).

Table 4.3 below shows that the mean duration of MDR TB patients in Gondar University Hospital were 28.834 months. The 95% confidence interval of mean duration of MDR TB patients treated at university of Gondar hospital lies between 26.712 and 30.956 months.

Table 4.3: over all mean for survival time of MDR TB patients and their 95%CI & SE.

Mean			
Estimate	Std. error	95% Confidence Interval	
		Lower Bound	Upper Bound
28.834	1.083	26.712	30.956

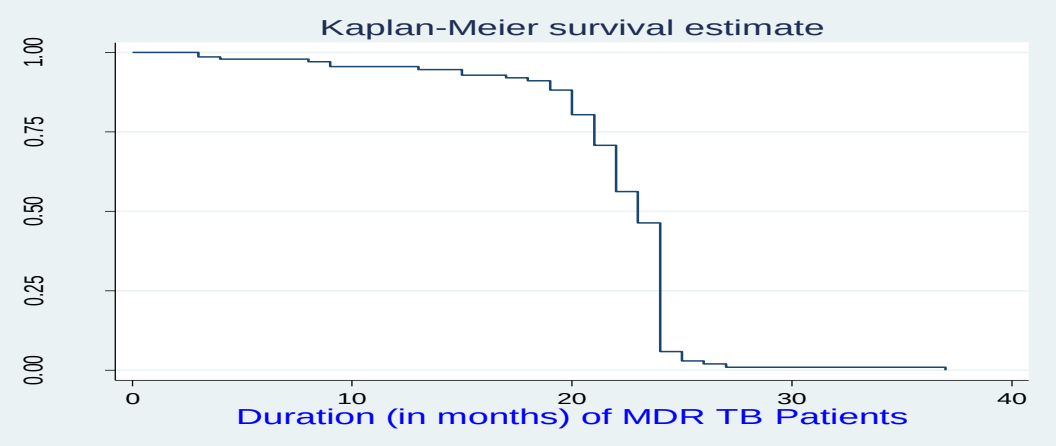


Figure 4.1: Overall product limit estimate of survival function

Table 4.4 shows that the average duration of male MDR TB patients (28.996 months) was higher than that of mean duration of (20.924months) of female sex. The mean duration of time to death was 23.753 months for MDR TB patients of younger age-group (18- 34 years old), 25.922 for age group of 35-54 years old and 19.070 months for older MDR TB patients. The average duration of time to death was higher in married (26.873 months) MDR TB patients and lower in

Widow/Widowed (18.667 months) MDR TB patients. The mean duration of time to death of single and divorced MDR TB patients were 23.459 months and 20.333 months respectively. Relatively the average duration of time to death of MDR TB patients was higher in patients who started treatment before one month after diagnosis (31.775 months) than who started treatment after one month of diagnosis (17.634 months). The mean duration of time to death of MDR TB patients was 33.408 months for patients who took more than INH and RIF and 17.547 months for patients who took only INH and RIF. Generally the average duration of time to death due to MDR TB was lower in non-addicted patients than addicted patients. The highest mean duration of time to death was 32.351 months for patients, who had no any clinical complication, 14.385 months for patients who had Pneumonia complication, 18.971 months for patients who had Pneumothorax complication, 18.800 months for patients who had Hemoptysis complication and 14.300 months for patients who had Cor pulmonal complication. The mean duration of time to death of HIV positive MDR TB patients was 11.147 months and HIV negative MDR TB patients was 31.216 months. The average duration of time to death was higher in non-chronic co-infected (32.004 months) MDR TB patients and lower in Myocardial infarction co-infected (8.333 months) MDR TB patients. The mean duration of time to death of Diabetes mellitus co-infected and Asthma MDR TB patients were 16.944 months and 12.778 months respectively.

Table 4.4: mean duration for survival time of MDR TB patients and their 95%CI & SE by different socio-demographic and clinical characteristics based on Kaplan-Meier technique

Covariates	Category	Mean duration of the patient	95% Confidence interval	
			Lower bound	Upper bound
Sex of the patient	Male	28.996	26.292	31.700
	Female	20.924	18.775	23.072
Age of the patient	18-34 years	23.753	22.647	24.859
	35- 54 Years	25.922	22.230	29.613
	>=55 years	19.070	15.861	22.273
Marital status of the patient	Single	23.459	21.283	25.636
	Married	26.873	23.654	30.092
	Separated/ Divorced	20.333	17.216	23.451
	Widow/Widowed	18.667	13.541	23.793

Educational label of the patient	Illiterate Read and Write Secondary Tertiary and above	22.042 19.764 30.656 19.425	19.067 17.662 26.934 16.003	25.016 21.865 34.375 22.847
Employment status of the patient	Employed Own Business Day laborer Unemployed	30.195 19.375 19.923 21.650	25.000 17.347 14.125 19.753	35.391 21.367 25.721 23.547
Therapeutic Delay	>= 1 month < 1 month	17.634 31.775	15.290 29.502	19.978 34.047
Number of first line drugs	INH and RIF only More than INH and RIF	17.547 33.408	15.488 31.282	19.606 35.533
Smoking status	Yes No	14.360 33.084	11.851 31.152	16.869 35.015
Alcohol use	Yes No	25.178 25.846	22.289 24.574	28.067 27.118
Any clinical complication	No complication Pneumonia Pneumothorax Hemoptysis Cor pulmonal	32.351 14.385 18.971 18.800 14.300	30.169 9.125 15.531 14.921 8.525	34.533 19.644 22.410 22.679 20.075
MDR Category	Previously Treated Previously not Treated	32.174 12.931	30.210 10.059	34.137 15.803
HIV Co-infection	Positive Negative	11.147 31.216	7.981 29.197	14.314 33.236
Presence of any chronic disease	No chronic disease Diabetes mellitus Myocardial infraction Asthma	32.004 16.944 8.333 12.778	29.983 12.954 4.876 7.823	34.024 20.935 11.790 17.732
Smear positivity	Positive Negative	29.337 18.720	27.021 15.963	31.653 21.477

Separate graphs of the estimates of the Kaplan-Meier functions for all covariates are presented in Figures 1C Appendix(C) in order to see whether there is difference in time to death between different categories of individuals. The Log-rank test was performed to investigate the significance of the observed difference in the Kaplan-Meier estimates of the survivor functions among different categories of the covariates. The Kaplan-Meier graphs for a covariate crossing

more than one times indicating that the variable is insignificant. All of the graphs show differences between different categories except sex of the patient, marital status of the patient, educational label of the patient, employment status of the patient, religion of the patient and smear positivity. The upper curve in each Figure indicates that the particular group experiences more survival time than the one below. In order to investigate if there is significant difference between the time of death of MDR TB patients by sex, Kaplan-Meier survivor estimates for the two gender groups are plotted in appendix C. This Figure shows that males had more survival time as compared to females. Similar analysis is performed to investigate differences in the time to death among the patients with respect to marital status, educational level, employed status, religion and smear positivity from the Kaplan-Meier curves in Appendix C shows that the curves cross each other more than one times indicating that there is no difference between time to death of these group patients. The result from Table 4.5 the p-value of the Log-rank test pointed out that all factors except sex of the patient, marital status of the patient, educational label of the patient, employment status of the patient, religion of the patient and smear positivity have differences in the survival experiences among their categories at 5% level of significance.

Table 4.5: Results of log-rank test of equality of survival distribution for the different categorical covariates

Covariates/Factors	Chi-square	Df	p-value
Sex of the patient	0.027	1	0.870
Age of the patient	15.231	2	0.000
Marital status of the patient	4.813	3	0.186
Educational label of the patient	1.555	3	0.670
Employment status of the patient	1.590	4	0.811
Religion of the patient	0.741	2	0.690
Therapeutic Delay	12.253	1	0.000
Number of first line drugs	21.260	1	0.000
Smoking status	48.971	1	0.000
Alcohol use	20.232	1	0.000
Any clinical complication	29.320	4	0.000
MDR Category	48.555	1	0.000
HIV Co-infection	50.166	1	0.000
Presence of any chronic disease	55.603	3	0.000
Smear positivity	1.272	1	0.259

+4.3 Cox Proportional Hazard Regression Model

4.3.1 Univar ate Cox proportional hazards model

After making a comparison of the survivorship experience among groups of covariates, the next important step is model development. An initial step in the model building process is to identify sets of explanatory variables that have the potential for being included in the linear components of a multivariable proportional hazards model. We started with fitting univariable Cox proportional hazards regression model. Table 2B (Appendix B) shows that fifteen univariable Cox proportional hazards models were fitted.

The result from Table 2B (Appendix B) shows that not all explanatory variables are important to fit multi-covariate Cox proportional hazards model. Thus, the important covariates were selected based on their contribution to the maximized log partial likelihood of the model (-$2LL_p(\widehat{\boldsymbol{\beta}})$). The value of -$2LL_p(\widehat{\boldsymbol{\beta}})$ for the null model is 401.960. Therefore, inclusions of covariates were based on the amount of reduction of this value and p-value. The bigger the reduction of this value, the better fit. Hence, based on the amount of reduction the log partial likelihood of variable smoking status has high contribution to maximized log partial likelihood of the model (-$2LL_p(\widehat{\boldsymbol{\beta}})$) because it shows highest reduction in (-$2LL_p(\widehat{\boldsymbol{\beta}})$). It reduces the value from 401.960 to 363.796. This reduction of 38.164 is highly significant (p-value < 0.0001) when compared with percentage points of the distribution on 1 degree of freedom. The second highest reduction in (-$2LL_p(\widehat{\boldsymbol{\beta}})$) is obtained from variable MDR category and it shows significant change. It reduces the value from 401.960 to 368.893, which reduced the statistic by 33.067. Then, the reduction in (-$2LL_p(\widehat{\boldsymbol{\beta}})$) due to the inclusion of chronic co-infection, HIV co-infection, alcohol intake, any clinical complication, number of drugs at initiation, age of the patient and therapeutic delay to the null model successively one at a time are 32.371, 29.007, 24.961, 23.604, 21.081, 17.621, 11.298 respectively. All of them are significant using the chi-square test (Wald test) at 5% level of significance. But the reduction in (-$2LL_p(\widehat{\boldsymbol{\beta}})$) due to the inclusion of marital status, educational label, religion of the patient, smear positivity, employment status and sex of the patient to the null model successively one at a time are 5.111, 1.567, 1.441, 1.15, 1.147 and 0.026 respectively and they are not significant at 5% level of significance. Hence all these covariates except marital status, educational label, religion of the patient, smear positivity,

employment status and sex of the patient are considered in the multivariable Cox regression model. The results are shown in table 3 of Appendix B. We then applied stepwise variable selection method (Collett, 2003) to obtain the reduced model.

At the next step of fitting the Cox proportional hazard model using the variables in Table 3B (Appendix B) and the value of (-2$LL_p(\hat{\boldsymbol{\beta}})$) will be 285.062. Then, omitting variables from the model will be based on the increasing in (-2$LL_p(\hat{\boldsymbol{\beta}})$) and p-value. The variable smoking status of the patient did not have significant increment and the p-value is 0.081. Therefore, the variable smoking status of the patient is excluded from the model and hence the reduced model consisted of the remaining six variables. Table 4.6 shows the fitted reduced model.

The final step in model development strategy is the consideration of interaction terms that may be useful in the improvement of the model. An interaction term is a new variable that is the product of two other variables in the reduced model. Note that there can be subject matter considerations that dictate that a particular interaction term (or terms) should be included in a given model, regardless of their statistical significance. In most settings there is no biological or clinical theory to justify automatic inclusion of interactions. The significance of each separate interaction is assessed by adding interaction terms to the main effects model one at a time and using the Wald test. All significant interactions should be included in the main-effects model. Then, examining the p-values of the Wald statistic in Table 4B (Appendix B) shows that there are no significant interaction effects at 5% level of significance.

Hence, our preliminary final model contains only the main effects of Therapeutic Delay, Number of drugs at initiation, Alcohol intake, any clinical complication, HIV co-infection and chronic co-infection. The following table shows the final model obtained from SPSS. But the interpretation based on this model should not be made until the adequacy of the proportional hazards Cox regression model has been checked.

Table 4.6: the Preliminary Final Model with parameter estimates and hazard ratios of the covariates

Covariates	B	SE	Wald X^2	Df	Sig.	Exp(B)	95% CI for Exp(B	
							Lower	upper
DELAY(>=1 month)	-1.265	0.425	8.874	1	0.003	0.282	0.123	0.649
NODRUG (INH &RIF only)	-1.434	0.414	12.026	1	0.001	0.238	0.106	0.536
Alcoholuse (Yes)	-1.570	0.629	6.241	1	0.012	0.208	0.061	0.713
(No Complication)			10.234	4	0.037			
Pneumonia	1.006	0.518	3.769	1	0.052	2.734	0.990	7.548
Pneumothorax	0.035	0.525	0.005	1	0.946	1.036	0.370	2.897
Hemoptysis	0.637	0.584	1.188	1	0.276	1.891	0.601	5.946
Cor pulmonal	1.726	0.568	9.233	1	0.002	5.615	1.845	17.090
HIV (Positive)	-1.488	0.385	14.939	1	0.000	0.226	0.106	0.480
Chronic (no chronic)			24.405	3	0.000			
Diabetes mellitus	1.283	0.504	6.479	1	0.011	3.608	1.343	9.691
Myocardial infraction	2.412	0.593	16.566	1	0.000	11.153	3.492	35.626
Asthma	1.699	0.474	12.862	1	0.004	5.469	2.161	13.841

The reference categories are those indicated in brackets

4.4 Model Adequacy

The adequacy of the model needs to be assessed after the model has been fitted to the observed survival data. At this point we have a preliminary fitted model and the next step is assessing the adequacy of the fitted model should be done in order to evaluate how well the fitted regression describes the data set. We started here first by checking the overall goodness of fit using R-square and LR, Score and Wald tests. We then proceed to check the proportionality assumption for each covariate included in the final model.

4.4.1 Overall Goodness of Fit

A perfectly adequate model may have low R^2 due to high percent of censored data. The value of the -2Log-Likelihood of the model with covariates in table 4.5 which is equal to 308.894 and the -2Log-Likelihood for the null or empty model equals 401.960. The measure of goodness of fit R^2_p is calculated as: R^2_p =1- exp[$\frac{2}{n}$(L_O - L_p)] = 1-exp[$\frac{2}{146}$((-200.98-(154.4470))] = 0.4713. Results of the Likelihood Ratio, Score and Wald tests for model goodness of fit displayed in Table 4.7 also suggests that model is good fit, i.e. significant at 5% level of significance.

Table 4.7: The Likelihood Ratio, Score and Wald tests for overall measures of goodness of fit of the preliminary final model in table 5A

Test	Chi-Square	Df	Pr > Chisq
Likelihood Ratio	93.0666	6	<.0001
Score	116.7047	6	<.0001
Wald	73.4573	6	<.0001

Testing Global Null Hypothesis: BETA=0

4.4.2 Testing the Proportional Hazards Assumption

There are two basic assumptions of the Cox Regression model. The first one is log-linearity, that is, the relationship between log hazard or log cumulative hazard and a covariate is linear. Since all covariates used in the final model are categorical, there is no need of checking linearity. The second one is proportional hazards (the time independence of the covariates in the hazard function, that is, the ratio of the hazard function for two individuals with different regression covariates does not vary with time). A proportional hazard is one of the very important assumptions in the Cox model.

The adequacy of the preliminary final model is checked for the validity of proportional hazards assumption using test based on the interaction between variables in the model with logarithm of survival time and assess their significance using the Wald test. Also the plot of the scaled schoenfeld residuals is used to provide any additional insight into any departure from proportionality. Under the assumption of proportionality of the proportional hazards model, the distribution of residuals over time is random and lowess smoothing line should be a straight line around zero, with no particular trend with time. Table 4.8 below shows the SAS output for testing proportionality assumption.

From Table 4.8 we can see the Wald chi-square values and the corresponding p-values for each covariate. Since the p-values for each interaction of covariate with logarithm of time are greater than 0.05, the proportionality assumption is satisfied. The global fit test also shows that the Wald chi-square test statistic is not significant which indicates that the proportional hazards assumption is not violated. Appendix C (Figures 1C to Figures 6C) shows the plots of the scaled Schoenfeld residuals for each covariate against time. The plot of the scaled schoenfeld Appendix C shows that the residuals are random without any systematic pattern and the

smoothed plot looks straight line without any departure from the horizontal line. Thus, there is no violation of proportional hazards assumption.

Table 4.8: Result of test of proportionality assumption for each covariate in the final model

Analysis of Maximum Likelihood Estimates						
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
DELAY	1	-1.06222	0.88614	1.4369	0.2306	0.346
NODRUG	1	-0.73675	0.91332	0.6507	0.4199	0.479
Alcoholuse	1	-1.00006	1.38314	0.5228	0.4697	0.368
Clnccomplicn	1	0.13676	0.28632	0.2282	0.6329	1.147
HIV	1	-0.59229	0.89481	0.4381	0.5080	0.553
CHRONIC	1	0.17826	0.34485	0.2672	0.6052	1.195
DELAYT	1	-0.21720	0.42376	0.2627	0.6083	0.805
NODRUGT	1	-0.27990	0.41979	0.4446	0.5049	0.756
AlcoholuseT	1	-0.28596	0.64064	0.1992	0.6553	0.751
ClnccomplicnT	1	0.04935	0.13591	0.1319	0.7165	1.051
HIVT	1	-0.43167	0.43439	0.9875	0.3204	0.649
CHRONICT	1	0.15722	0.16481	0.9100	0.3401	1.170

Linear Hypotheses Testing Results			
Label	Wald Chi-Square	DF	Pr > ChiSq
test_proportionality	4.3879	6	0.6243

4.5 Interpretation and discussion of the results

The study assessed survival of MDR TB patients and examined the socio-demographic and clinical determinants of MDR TB patients in Gondar university hospital. In survival analysis the measure of effect is the hazard ratio. It is interpreted in the same way as the odds ratio. The higher the hazard ratio the lower is the survival probability, and vice versa. If for an exposed group the hazard ratio is high, the survival probability would be equivalently low. From the final model in Table 4.6 we obtained six significant main effects: Therapeutic delay, number of drugs at initiation, alcohol intake, any clinical complication, HIV co-infection and chronic co-infection of the patients.

Therapeutic delay has a significant negative association with mortality of MDR TB patients. After adjusting other covariates, the estimated coefficient is -1.265 for a confirmed MDR-TB

patient who starts treatment before one month and the hazard ratio is 0.282. This indicates that the hazard of death of MDR TB patients is reduced by 71.8 % for a confirmed MDR TB patient who starts treatment before one month relative to a confirmed MDR TB patients who started treatment after one month (adjusted HR=0.282, 95% CI=0.123-0.649). The 95% confidence intervals also suggest that the rates could be as low as 0.123 and as high as 0.649. that means the risk of death of MDR TB patients who starts MDR treatment before one month could be as low as 0.123 and as high as 0.649. This finding is consistent with Theodros Getachew *et al*. (2013).

The result of this study revealed that a negative relationship between number of drugs taken at initiation and duration of MDR TB patients. After adjusting for other covariates, the estimated coefficient is -1.434 for MDR TB patients who took more than INH and RIF at initiation and the hazard ratio is 0.238. This indicates that the hazard of death of MDR TB patients is reduced by 76.2 % for a confirmed MDR TB patient who took more than INH and RIF at initiation relative to a confirmed MDR TB patients who takes INH and RIF only (adjusted HR=0.238, 95% CI=0.106-0.536). The study revealed that the risk of death of MDR TB patients who takes more than INH and RIF drugs at initiation is lower than patients who take only two drugs (INH and RIF only) at initiation.

The hazard ratio of MDR TB patients who are not using alcohol as compared to those who use alcohol is 0.208. This means the risk of death for MDR TB patients who are not using alcohol is about 0.208 times higher than MDR TB patients who are used alcohol. The 95 % confidence interval also suggests that the risk of death for MDR TB patient who are not used alcohol is 0.061 times as low and 0.713 times as large as those who are used alcohol drink. Alcohol use is negatively related with MDR TB patient's survival chances. A similar trend that MDR-TB subjects have an overall lower lifetime prevalence of any alcohol use than the non-alcohol user was also observed in some other study in Botswana Zetola N. *et al*. (2012).

After adjusting for other covariates, the hazard of death of MDR TB patients with Pneumonia complication is 2.734 times higher than MDR TB patients who have no any clinical complication (adjusted HR=2.734, 95% CI: 0.990, 7.548). The hazard of death of MDR TB patients with Pneumothorax complication is 1.036 times higher than MDR TB patients who have no any clinical complication (adjusted HR=1.036, 95% CI: 0.370, 2.897). The hazard of death of MDR TB patients with Hemoptysis complication is 1.891 times higher than MDR TB patients who

have no any clinical complication ((adjusted HR=1.891, 95% CI: 0.601, 5.946). The hazard of death of MDR TB patients with Cor pulmonal complication is 5.615 times higher than MDR TB patients who have no any clinical complication (adjusted HR=5.615, 95% CI: 1.845, 17.090). All these hazard ratios indicate that the risk of death of MDR TB patients with different clinical complication is higher than relative to MDR TB patients with no clinical complication.

The estimated relative risk (hazard ratio) of time to death for MDR TB patients who are HIV negative as compared to HIV positive patient is 0.226 (95% CI: 0.106-0.480). This indicates that the hazard of death of MDR TB patients is reduced by 77.4 % for confirmed MDR TB patients who are HIV negative relative to a confirmed MDR TB patients who are HIV positive. The 95% confidence interval also suggests that the risk of death for HIV negative MDR TB patients could be as low as 0.106 and as high as 0.480. Hence, HIV co-infected MDR TB patients have a relatively-shorter duration than HIV negative MDR TB patients. This finding is consistent with Yanina Balabanova *et al*. (2011) and Samuel OM Manda *et al*. (2004).

After adjusting other covariates, the hazard of death of MDR-TB patients with Diabetes mellitus co-infection is 3.608 times higher than MDR-TB patients who have no chronic co-infection (adjusted HR=3.608, 95% CI: 1.343, 9.691). The hazard of death of MDR-TB patients with myocardial infarction co-infection is 11.153 times higher than MDR TB patients who have no chronic co-infection (adjusted HR=11.153, 95% CI: 3.492, 35.626). The hazard of death of MDR- TB patients with Asthma co-infection is 5.469 times higher than MDR-TB patients who have no chronic co-infection (adjusted HR=5.469, 95% CI: 2.161, 13.841). These all hazard ratios indicate that the risk of death of MDR-TB patients with different chronic co-infection is higher than relative to MDR-TB patients with no chronic co-infection. This finding is consistent with Matthew J. Magee *et al*. (2014) and Kang YA *et al*. (2013)

4.6. Parametric Regression Modelling of Survival Time of MDR TB Patients

4.6.1. Model Selection for Survival Time of MDR TB Patients

For the survival of MDR tuberculosis patients the parametric regression models were fitted. We consider model comparison after adjusting for the effect of covariates. In this case the graphical displays are based on the Cox-Snell plots. That is, if the model is good, the plot of Cox-Snell residuals versus Nelson-Aalen cumulative hazard estimates should lie along the 45 degree diagonal line that passes through the origin. Using all the covariates in the study, we fitted two

parametric regression models the Exponential and Weibull models with the corresponding AIC and BIC values. Here we present the Cox-Snell plots for model comparison in Figures 4.2 to 4.3.

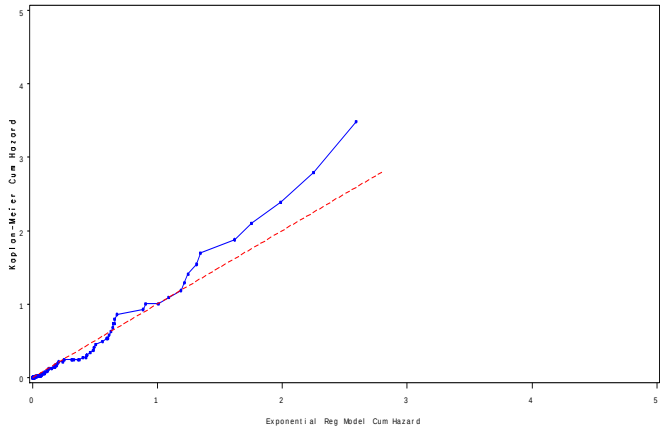


Figure 4. 2 The Cox Snell plot after fitting Exponential regression mode

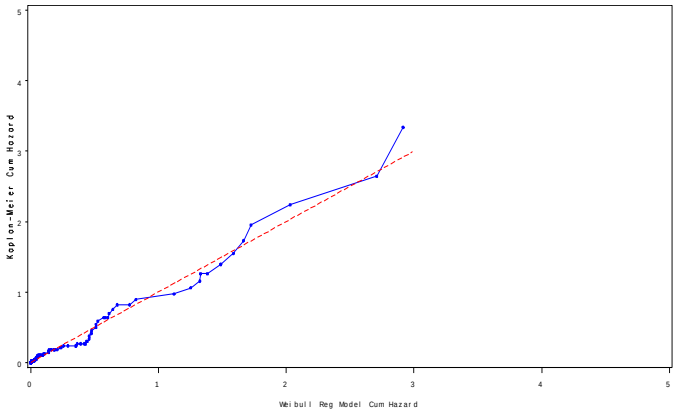


Figure 4. 3 The Cox Snell plot after fitting Weibull regression model

From the Cox-Snell plots in Figures 4.2 and 4.3, it is observed that Weibull regression model seems the best fit among the two models. But graphical methods may not assure the result. The common applicable criterion to select the model is the Akaikie information criterion (AIC) statistic proposed by Akaikie (1974). So, In addition to the graphical comparison of the three parametric regression models, I used Akaikie information criterion (AIC) and Bayesian information criterion (BIC) to choose the best model out of the four possible models. The STATA output of the two parametric survival regression models are displayed in appendix B from table 5 to table 6 with the corresponding AIC and BIC values.

Table 4.9: Statistics for model comparison

Model	**Observation**	**ll (null)**	**Ll (model)**	**AIC value**	**BIC value**	**DF**
Exponential	146	-132.3478	-77.32542	186.6508	234.3886	16
Weibull	146	-132.345	-72.65042	179.3008	230.0222	17

According to the results in Table 4.9, the Weibull regression model with the smallest value of AIC and BIC seems to be the best fit of the two models. Nevertheless, the results of cox-snell were consistent with the results based on Akaikie‟s information criterion. Thus, the Weibull regression model is considered further to discuss the effect of covariates on the survival of MDR TB patients.

4.6.2 Univariate unadjusted Weibull regression model

As Weibull regression is selected, According to the Weibull analysis of single covariate, the selected risk factors for further analysis and interpretation are made here below. To have an idea about the individual effects of the different explanatory variables on survival of MDR TB patients, we fitted Weibull regression model separately for each explanatory covariate. The results are shown in Table 4.10 bellow.

Table 4.10.The result of un-adjusted univariate analysis using Weibull regression model

covariate	Estimate	Std.error	Chi-square	p-value	Df	-2*LL	95% CI	
							Lower	Upper
Sex	0.0549	0.3178	0.03	0.8628	1	264.6605	-0.5680	0.6779
Age	-0.6862	0.2177	9.94	0.0016	2	252.1908	-1.1129	-0.2595
MRSUS	-0.2634	0.1739	2.30	0.1297	3	262.427	-0.6042	0.0773
EDULABL	0.0488	0.1615	0.09	0.7627	3	264.598	-0.2678	0.3653
EMPSUS	0.0157	0.1051	0.02	0.8809	4	264.667	-0.1902	0.2217
Religion	0.2978	0.5298	0.32	0.5740	2	264.327	-0.7405	1.3361
DELAY	1.0588	0.3389	9.76	0.0018	1	252.809	0.3945	1.7231
NODRUG	1.4989	0.3873	14.97	0.0001	1	241.482	0.7397	2.2580
Smokstatus	1.8698	0.3771	24.58	<0.0001	1	223.345	1.1307	2.6090
Alcoholuse	2.1721	0.6565	10.95	0.0009	1	239.364	0.8853	3.4589
Clnccomplicn	-0.3829	0.1092	12.29	0.0005	4	251.054	-0.5970	-0.1688
MDRCAT	-1.7724	0.3515	25.42	<0.0001	1	229.927	-2.4614	-1.0834
HIV	1.8464	0.3429	28.99	<0.0001	1	232.998	1.1742	2.5185
CHRONIC	-0.7030	0.1398	25.28	<0.0001	3	235.643	-0.9771	-0.4290
SMEAR	-0.4034	0.3702	1.19	0.2759	1	263.560	-1.1289	0.3222

Results displayed in Table 4.10, illustrate that the statistically significant risk factors for the survival probability of multidrug resistant tuberculosis patients are Age of the patient, Therapeutic delay, Number of first line drugs at initiation, Smoking status, Alcohol use, any clinical complication, MDR category, HIV co-infection and chronic co-infection. Whereas the risk factors that were not statistically significant are sex of the patient, marital status, educational label of the patient, employment status, religion of the patient and smear positivity at 5% level of

significance. The risk factors those were statistically significant included in the final Weibull regression model for the prediction of survival probability of MDR TB patients.

4.6.3 Multivariable Analysis Weibull regression model

When there are a number of explanatory variables of possible relevance, the effect of each term cannot be studied independently of the others. The effect of any given term therefore depends on the other terms currently included in the model. However, in the univariate analysis technique the relations that are obtained for one factor do not take into account the other factors. So the multivariable analysis is used to know the most important factors associated with mortality of MDR TB patients in relation to the covariates included in the model. After fitting the univariate weibull survival regression analysis the next step is selecting the most important variables to fit the multivariate weibull regression model. In order to select the most important covariates in the final model, we used stepwise variable selection. The results of the stepwise selection are displayed in table 8 of appendix B.

Results presented in Table 4.11 indicate the parameter estimates of coefficients for the covariates in the final Weibull regression model along with the associated significance level, hazard ratio with corresponding standard error and 95% confidence interval for the hazard ratio. Survival time of MDR TB patients were significantly related with age of the patient, therapeutic delay, number of drugs at initiation, alcohol use, any clinical complication, MDR category, HIV co-infection and chronic co-infection as can be seen from the Table 4.11 bellow.

Table 4.11.Summary of parameter estimates of the final multivariate weibull model

Covariates	Coef.	Std. Err.	z	P>|z|	Haz. Ratio	[95% Conf. Interval] for Haz. Ratio	
						Lower	Upper
Age (18-34 years)							
35-54 years	1.050358	1.584558	1.89	0.058	2.858673	0.9645951	8.471963
>=55 years	1.460258	2.513778	2.50	0.012	4.307071	1.372104	13.52001
DELAY (>=month)	-1.382157	0.1070241	-3.24	0.001	0.2510366	0.1088542	0.5789337
NODRUG (INH &RIF only)	-1.288478	0.1144644	-3.10	0.002	0.27569	0.1221824	0.6220619
Alcoholuse (Yes)	-1.441549	0.1481942	-2.30	0.021	0.236561	0.0692957	0.8075703

Clnccomplicn (No)							
Pneumonia	0.418614	0.861957	0.74	0.460	1.519854	0.5001017	4.618974
Pneumothorax	0.292495	0.758625	0.66	0.509	1.422056	0.4998341	4.045831
Hemoptysis	0.292495	0.794574	0.49	0.622	1.339767	0.4189941	4.284011
Cor pulmonal	1.700102	3.187495	2.92	0.004	5.474506	1.748793	17.13766
MDRCAT (Previously treat)	0.7966434	0.9148459	1.93	0.053	2.218083	.9883271	4.978001
HIV (Positive)	-1.277784	0.1155582	-3.08	0.002	0.278654	0.123615	0.6281446
Chronic (no chronic)							
Diabetes mellitus	1.209424	1.792501	2.26	0.024	3.351552	1.174898	9.560745
Myocardial infraction	1.796817	3.892995	2.78	0.005	6.030419	1.701556	21.37218
Asthma	1.689863	2.677192	3.42	0.001	5.418741	2.057559	14.27067
ln_p	0.4955398	0.1329137	3.73	0.000	0.4955398	0.2350337	0.7560459
P	1.641384	0.2181624			1.641384	1.264951	2.129838
1/p	0.6092419	0.0809766			.6092419	.4695193	.7905442

The reference categories are those indicated in brackets.

4.6. 4 Assessment of Adequacy of the Weibull Regression Model

The likelihood ratio test presented in Table 4.12, it illustrate that the model was significantly fit the data of MDR TB patients and in using the log likelihood values of the null model and the full model, it can be seen that the model has a significant improvement after the covariate is incorporated in the model.

Table 4.12 The likelihood ratio and significance of the Weibull regression model

Log likelihood (intercept only)	Log likelihood (Model)	LR chi-square	DF	Prob > chi2	Intercept	Scale
-132.34495	-75.385077	113.92	8	0.0000	-4.114011	1.529221

4.6.5. Interpretation and Discussion of the Weibull Regression Model

From the Weibull regression model, age of the patients is positively associated with mortality of MDR TB patients. After adjusting the other covariates, the hazard rate of MDR TB patients who had age group 35 -54 years was 2.858673 times than of MDR TB patients who had age group

18-34 years (adjusted HR=2.858673,95% CI=0.9645951,8.471963). The hazard rate of MDR-TB patients who had age group are >=55 years was 4.307071 times than of MDR TB patients who had age group 18-34 years (adjusted HR=4.307071,95% CI=1.372104 , 13.52001). Generally the finding of this study revealed that an increase of age of the patients declines the survival probability of MDR-TB patients. Some other studies showed that age is not a significant factor for mortality of MDR-TB patients Getachew *et al*. (2013) and Ahmed M. Salih *et al*. (2010)

For therapeutic delay, the study revealed that therapeutic delay is a negative significant factor for mortality of MDR-TB patients and the hazard rate of MDR TB patients who started treatment before one month delay after diagnosis is 0.2510366 times that of MDR-TB patients started treatment after one month of diagnosis (adjusted HR = 0.2510366, 95% CI: 0.1088542, 0.5789337). This indicates that the risk of death of MDR-TB patients who started treatment before one month delay after diagnosis is declined by 74.9%. Adjusting the other covariates, the hazard rate of MDR-TB patients who have taken more than INH and RIF drugs at initiation is 0.27569 times than MDR-TB patients who have taken only INH and RIF at initiation (adjusted HR = 0.27569 , 95% CI: 0.1221824 , 0.6220619). This result indicates that the probability of survival of MDR-TB patients who had taken more than INH and RIF drugs at initiation is higher than relative to MDR-TB patients who had taken only INH and RIF drugs at initiation. Number of drugs taken at initiation is negatively significant risk factor for mortality of MDR-TB patients at Gondar university hospital. These results confirm the result obtained from the previous studies (Theodros Getachew *et al*. (2013)).

The addiction of alcohol is another prognostic factor that significantly predicts the mortality of MDR-TB patients. The result obtained from this study indicates that the hazard ratio of non-alcohol takers is 0.236561 times that of alcohol takers (adjusted HR = 0.236561, 95% CI: 0.0692957, 0.8075703). There is a significantly negative relationship between alcohol use and mortality of MDR-TB patients. This result confirms that MDR-TB subjects have an overall lower lifetime prevalence of any alcohol use than the non -alcohol user N. M. Zetola *et al*. (2012).

After adjusting other covariates, the hazard of death of MDR-TB patients with Pneumonia complication is 1.519854 times higher than MDR-TB patients who have no any clinical

complication (adjusted HR=1.519854 , 95% CI: 0.5001017,4.618974). This result revealed that the risk of death of MDR TB patients with Pneumonia complication is 51.9% higher than that of MDR TB patients have no any clinical complications. The hazard of death of MDR-TB patients with Pneumothorax complication is 1.422056 times higher than MDR-TB patients who have no any clinical complication (adjusted HR=1.422056, 95% CI: 0.4998341, 4.045831). This result indicates that the risk of death of MDR-TB patients with Pneumothorax complication is 42.2% higher than that of MDR-TB patients have no any clinical complications. The hazard of death of MDR-TB patients with Hemoptysis complication is 1.339767 times higher than that of MDR-TB patients who have no any clinical complication ((adjusted HR=1.339767,95% CI: 0.4189941 , 4.284011). This result revealed that the risk of death of MDR-TB patients with Hemoptysis complication is 33.9% higher than that of MDR-TB patients have no any clinical complications. The hazard of death of MDR-TB patients with Cor pulmonal complication is 5.474506 times higher than MDR TB patients who have no any clinical complication (adjusted HR=5.474506 , 95% CI: 1.748793, 17.13766). These all hazard ratios indicate that the risk of death of MDR-TB patients with different clinical complication is higher than relative to MDR-TB patients with no clinical complication.

The estimated relative risk (hazard ratio) of time to death for MDR-TB patients who are previously not treated as compared to previously treated MDR-TB patients is 2.218083 (95% CI: 0.9883271, 4.978001). This indicates that the hazard of death of MDR TB patients is higher for confirmed MDR-TB patients who are previously not treated relative to a confirmed MDR-TB patient who are previously treated. The 95% confidence interval also suggests that the risk of death for previously not treated MDR-TB patients could be as low as 0.9883271and as high as 4.978001. Hence, MDR-TB patients who are previously not treated have a relatively-shorter duration than previously treated MDR-TB patients. This finding is consistent with E.E Telzak *et al*. (1998) and Songhua Chen *et al*. (2013).

The estimated relative risk (hazard ratio) of time to death for MDR-TB patients who are HIV negative as compared to HIV positive patient is 0.278654 (95% CI: 0.123615, 0.6281446). This indicates that the hazard of death of MDR-TB patients is reduced by 62.2 % for confirmed MDR-TB patients who are HIV negative relative to a confirmed MDR-TB patients who are HIV positive. The 95% confidence interval also suggests that the risk of death for HIV negative

MDR-TB patients could be as low as 0.123615 and as high as 0.6281446. Hence, HIV co-infected MDR-TB patients have a relatively-shorter duration than HIV negative MDR-TB patients. This finding is consistent with Yanina Balabanova *et al*. (2011) and Samuel OM Manda *et al*. (2004).

After adjusting other covariates, the hazard of death of MDR-TB patients with Diabetes mellitus co-infection is 3.351552 times higher than MDR-TB patients who have no chronic co-infection (adjusted HR=3.351552, 95% CI: 1.174898 ,9.560745). The hazard of death of MDR-TB patients with myocardial infarction co-infection is 6.030419 times higher than MDR-TB patients who have no chronic co-infection (adjusted HR=6.030419 ,95% CI: 1.701556 , 21.37218). The hazard of death of MDR-TB patients with Asthma co-infection is 5.418741 times higher than MDR-TB patients who have no chronic co-infection (adjusted HR=5.418741, 95% CI: 2.057559, 14.27067). These all hazard ratios indicate that the risk of death of MDR-TB patients with different chronic co-infection is higher than relative to MDR-TB patients with no chronic co-infection. This finding is consistent with Matthew J. Magee *et al*. (2014) and Kang YA *et al*. (2013).

CHAPTER FIVE

5. CONCLUSION AND RECOMMENDATION

5.1 Conclusion

- This study was aimed to identify the survival and predictors of mortality among patients under Multi-Drug Resistant Tuberculosis treatment at Gondar University teaching Hospital. Fifteen covariates were selected for the study for determining the risk factors of mortality of MDR TB patients and modeling the survival time, a total of 146 patients were included in the study out of which 28.7% were died and the rest 71.3% were censored. 15 univariable Cox Proportional Hazards regression Model and Weibull regression model were developed to assess the relation between the survival status of MDR TB patients and the selected variables. Based on the results, the multi-variable Cox Proportional Hazards regression Model and Multivariate Weibull regression model of duration of survival status was employed to select the most important determinants and the research has shown clinical factors are critical in determining survival of patients under MDR-TB treatment than socio-demographic factors.
- The Cox regression analysis showed that Therapeutic delay, initial number of drugs at initiation , alcohol use, clinical complication, HIV status of patient and chronic co-infection were the major factors that affect the survival probability of MDR TB patients at Gondar university hospital. In the other hand it was found that factors which had no significant impact on the survival of MDR TB patients were sex of the patient, age, marital status, educational level, employment status, religion, smoking status, MDR category and smear positivity of the patients. Higher hazard of death or lower survival rate was noted in patients who started treatment after a month of period diagnosed as MDR-TB, patients who take only INH and RIF only, patients who drink alcohol, patients who have clinical complications during the treatment period, HIV positive and patients who have chronic co-infection.
- The two parametric regression models: Exponential and Weibull regression models, for survival probability of MDR TB patients were compared. The Weibull regression model was found to better fit to the data. The major factors that predict the survival probability of TB patients were age of the patient, therapeutic delay, number of drugs taken at

initiation, alcohol use, any clinical complication, MDR category, HIV co-infection and chronic co-infection. Elderly MDR TB patients had higher hazard rate and higher hazard of death or lower survival time was noted in patients who started treatment after a month of period diagnosed as MDR-TB. However, Patients who took more than INH and RIF at initiation, patients who are HIV negative, patients who are previously treated and patients who did not take alcohol have higher survival probability (lower hazard rate). MDR`-TB patients who have clinical complication and chronic co-infection during treatment have lower survival probability (higher hazard rate) relative to MDR-TB patients who have no any clinical complication and no chronic co-infection at Gondar university hospital.

5.2 Recommendation

Based on the result of the study different factors are identified for the mortality of multidrug resistant tuberculosis patients. The following recommendations are made for policy makers, clinicians and the public at large.

- Awareness have to be given for the society on the risk factors of MDR-TB and taking care for starting treatment before one month delay after diagnosis to improve their survival time.
- MDR-TB patients who took only INH and RIF have high risk of death. So, patients have to come on time to the treatment center and took more than INH and RIF drugs.
- Elderly MDR-TB patients and previously not treated patients have high hazard of death. So that special attention should be given for elderly MDR-TB patients. In addition tuberculosis patients should be treated on time to minimize and eliminate the development of MDR-TB.
- This study shows that main predictive factors for the survival time of MDR-TB patients are more clinical variables, so health workers and stakeholders should be cautious when patients are alcohol users, have any clinical complications and chronic co-infections, HIV positive and they took only INH and RIF at initiation.
- The government and concerned bodies should work on perception about the problem and its risk factors of MDR TB, so that patients should be well informed about the problem and early diagnose to give treatment and to stop mortality due to MDR TB.

REFERENCES

1. Ahmad M. Salih, Muayad A. Merza (2010). Risk factors for multi-drug resistant tuberculosis: a review Duhok Med J, **4** (2):1-7.
2. Akaike, H. (1974). A new look at the statistical model identification", IEEE Transactions on Automatic Control, **19** (6): 716–723, doi:10.1109/TAC.1974.1100705, MR 0423716
3. Alena, S. Sven, H. Valiantsin, R. *et al.* (2013). Multidrug-resistant tuberculosis in Belarus: the size of the problem and associated risk factors: Bull World Health Organ, 91:36–45
4. Balabanova, Y, Radiulyte, B, Davidaviciene, E. *et al*. (2011). Survival of drug resistant tuberculosis patients in Lithuania: retrospective national cohort study. BMJ Open, **1**: e000351. doi:10.1136/ bmjopen-2011-000351
5. Bijan, M., Safaee, A., Pourhoseingholi, M., Fatemi, R., Tabeie, Z., Zali, M. (2008). Statistical Comparison of Survival Models for Analysis of Cancer Data. Asian Pacific Journal of Cancer Prevention, Vol **9**, 2008
6. Chest Clinic, Blackburn Royal Infirmary, Blackburn, Lancs BB2 3LR, and Postgraduate School of Medicine and Health, University of Central Lancashire, Preston, Lancs PR1 2HE, UK (2005). Multidrug-resistant tuberculosis (MDR-TB): epidemiology, prevention and treatment; British Medical Bulletin, 73 and 74: 17–24
7. Collett, D. (2003). Modeling survival data in medical research: Texts in Statistical Science (2nd edition.). Chapman & Hall/CRC.
8. Cox, H., Kalon, S., Allamuratova, S., Sizaire,V., Tigay, Z. *et al*.(2007). Multidrug-Resistant Tuberculosis Treatment Outcomes in Karakalpakstan, Uzbekistan: Treatment Complexity and XDR-TB among Treatment Failures. PLoS ONE 2(11).
9. Daniel, O., Oladapo, O. and Alausa, O. (2006). Default from tuberculosis treatment programme in Sagamu, Nigeria. Niger J Med, 15:63–67.
10. Dhingra, K. Rajpal, S., Anshu M. and Hanif, M. (2008). Outcome of multi-drug resistant tuberculosis cases treated by individualized regimens at a tertiary level clinic; Indian J Tuberc; 55:15-21
11. Dias Baptista I. M. F.; Usó S. M. R. S.; Marcondes-Machado J (2008). Trends in multidrug-resistant tuberculosis, J. Venom. Anim. Toxins incl. Trop. Dis vol.14 no.2 Botucatu.

12. Documents WHO/HTM/TB/2004.330 and WHO/HTM/HIV/2004.1
13. Doo Soo Jeon, Dong Ok Shin, Seung Kyu Park *et al*. (2011). Treatment Outcome and Mortality among Patients with Multidrug-resistant Tuberculosis in Tuberculosis Hospitals of the Public Sector: The Korean Academy of Medical Sciences; 26: 33-41
14. Drobniewski F, Eltringham I, Graham C, Magee J G, Smith E G, Watt B (2002). A national study of clinical and laboratory factors affecting the survival of patients with multiple drug resistant tuberculosis in the UK; Thorax, 57:810–816;.
15. Emerging Infectious Diseases • www.cdc.gov/eid • Vol. 14, No. 9, September 2008. Underreported Threat of Multidrug- Resistant Tuberculosis in Africa Yanis Ben Amor, Bennett Nemser, Angad Singh, Alyssa Sankin, and Neil Schluger.
16. Espinal M.A., Laserson K., Camacho M., Fusheng Z., Kim S.J. and Tlali R.E. *et al.* (2001) Determinants of drug-resistant tuberculosis: analysis of 11 countries, Int J Tuberc *Lung Dis* 5, 885–886.
17. Federal democratic republic of Ethiopia, Ministry of health, Health Sector Development Programme IV 2010/11 – 2014/15
18. Federal democratic republic of Ethiopia, Ministry of health. HSDP IV, annual performance report 2010/2011; 2010.
19. Federal ministry of health of Ethiopia: guideline for program and clinical management of drug resistant tuberculosis First edition; April, 2009
20. Federal ministry of health of Ethiopia: guidelines on programmatic management of drug resistant tuberculosis in Ethiopia second edition; December, 2013
21. Federal ministry of health of Ethiopia: on programmatic management of drug resistant tuberculosis in Ethiopia: training for general health workers; participant's manual second edition; December, 2013
22. Frieden, T., Sterling, T., Pablos-Mendez, A., Cauthen G M, Dooley S W. The emergence of drug-resistant tuberculosis in New York City. N Enjl J Med 1993; 328:521-526.
23. Guideline for the programmatic management of drug resistant tuberculosis: Emergency update 2008 (WHO/HTM/TB/2008.402). Geneva: World Health Organization.
24. Hosmer and Lemeshow (1989). Applied survival analysis. New York, John Wiley and sons.

25. Hosmer D, Lemeshow S. (1999). Regression modeling of time-to-event data. Wiley, New York.

26. Iseman M. (1993). Treatment of multi-drug resistant tuberculosis. N Engl J Med, 329, 784–91.

27. James, C., Brust, M., *et. a l*(2010). High Treatment Failure and Default Rates for Patients with MDR TB in KwaZulu-Natal, South Africa, 2000–2003; Int J Tuberc Lung Dis, 14(4): 413–419

28. Kang Y.*et al*. (2013). Impact of diabetes on treatment outcomes and long-term survival in multidrug-resistant tuberculosis.

29. Matthew J. Magee, Russell R. Kempker, Maia Kipiani *et al.* (2014). Diabetes Mellitus, Smoking Status, and Rate of Sputum Culture Conversion in Patients with Multidrug-Resistant Tuberculosis: A Cohort Study from the Country of Georgia; PLoS ONE 9(4): e94890. doi:10.1371/journal.pone.0094890

30. N. Zetola, C. Modongo, E. C. Kip, R. Gross, G. P. Bisson, and R. G. Collman (2012). Alcohol use and abuse among patients with multidrug-resistant tuberculosis in Botswana; Int J Tuberc Lung Dis, 16(11): 1529–1534. doi:10.5588/ijtld.12.0026.

31. Nihal, A., and Tekin, S. (2007). Cox regression models with non-proportional hazards applied to Lung cancer survival data. Hacettepe Journal of Mathematics and Statistics, 36 (2), 157- 167.

32. Park M M, Davis A L, Schluger N W, Cohen H, Rom W N. outcome of MDR-TB patients, 1983-1993: prolonged survival with appropriate therapy. Am J Crit care Med 1996; 153:317-324.

33. Samuel, M. , Lieketseng, J., Joey, L. and Martie, L.(2013). Risk of Death among HIV Co-Infected Multidrug Resistant Tuberculosis Patients, Compared To Mortality in the General Population of South Africa; J AIDS Clinic Res S3: 007. doi:10.4172/2155-6113.S3-007

34. Selamawit, H., Girmay, M. and Belaineh, G.*et al.* (2013). Determinants of multi-drug resistant tuberculosis in patients who underwent first-line treatment in Addis Ababa: a case control study: BMC Public Health 2013, 13:782

35. Seyed, M., Payam T. and Kaveh K., *et al.* (2005). Treatment of multiple drug-resistant tuberculosis (MDR-TB) in Iran; International Journal of Infectious Diseases 9, 317-322;.

36. Songhua, C. , Pengcheng, H. and Xiaomeng, W. (2013): Risk factors for multi-drug resistance among previously treated patients with tuberculosis in eastern China: a case–control study; International Journal of Infectious Diseases 17 (2013) e1116–e1120
37. Telzak, E. , Chirgwin, K. , Nelson, E *et al*. (1998). Predictors of multidrug-resistant tuberculosis among HIV-infected patients and response to specific drug regimens; INT J Tuber clung DIS 3 (4): 337-343
38. The world health report 2006: working together for health. Geneva, World Health Organization, 2006.
39. Theodros, G., Alemayehu, B. and Berhe, W. (2013). Survival and predictors of mortality among patients under multi-drug resistant tuberculosis treatment in Ethiopia: st. peter's specialized tuberculosis hospital, Ethiopia: IJPSR, Vol. **4** (2): 776-787
40. Turett, S., Telzk, E. and Torian, L. (1995). Improved outcomes for patients with multidrug-resistant tuberculosis. Clin Infect Dis; 21: 1238-1244.
41. Weibull, W. (1939). A Statistical Theory of the Strength of Materials. Ingenioers vetenskaps akakemien Handlingar, 151, 29 3-297.
42. Weibull, W. (1951). A Statistical Distribution of Wide Applicability. Journal of Applied Mathematics, 18, 29 3-297.
43. WHO (1996) Guidelines on the Management of Drug-Resistant Tuberculosis. WHO/TB/96.210. Geneva
44. WHO: Prevention and control of multidrug-resistant tuberculosis and extensively drug-resistant tuberculosis Report by the Secretariat; Document WHO/HTM/TB/2008/2009.394.
45. World Health Organization (2010). Multidrug and extensively drug-resistant tb (m/xdr-tb) 2010 global report on surveillance and response
46. World Health Organization. Global tuberculosis control (2008). Surveillance, planning, financing: http://www.who.int/tb/publications/global_report/2008/pdf/fullreport.pdf. Accessed 2008 Jul 24.
47. World Health Organization. Global Tuberculosis Control: who report; 2013.
48. World health organization. Tuberculosis MDR-TB & XDR-TB progress report; 2011.
49. Yashodhara, M., *et al.* (2010). Multidrug and extensively drug-resistant tuberculosis from a general practice perspective: Infection and Drug Resistance, 3 115–122.

Appendix B

Table 1B Kaplan-Meier estimates of time to death for MDR TB patients

Time	Total number of patients at risk	Total number of Deaths	Total number of Censored	Survival	Std. Error	[95% Conf. Int.]	
						Lower bound	Upper bound
1	146	3	0	0.9795	0.0117	0.9377	0.9933
2	143	2	0	0.9658	0.0151	0.9197	0.9856
3	141	3	2	0.9451	0.0189	0.8931	0.9721
4	136	2	1	0.9311	0.0210	0.8757	0.9623
5	133	4	0	0.9031	0.0246	0.8419	0.9414
6	129	3	0	0.8821	0.0269	0.8172	0.9250
7	126	1	0	0.8751	0.0275	0.8091	0.9194
8	125	0	1	0.8751	0.0275	0.8091	0.9194
9	124	3	2	0.8538	0.0295	0.7845	0.9021
10	119	1	0	0.8466	0.0301	0.7764	0.8962
11	118	2	0	0.8322	0.0313	0.7602	0.8843
12	116	4	0	0.8035	0.0333	0.7282	0.8600
13	112	2	1	0.7891	0.0342	0.7124	0.8476
14	109	2	0	0.7746	0.0351	0.6966	0.8350
15	107	2	2	0.7600	0.0359	0.6807	0.8222
16	103	2	0	0.7453	0.0367	0.6647	0.8092
17	101	1	1	0.7379	0.0371	0.6568	0.8027
18	99	2	1	0.7229	0.0378	0.6407	0.7893
19	96	2	3	0.7076	0.0385	0.6244	0.7757
20	91	0	8	0.7076	0.0385	0.6244	0.7757
21	83	0	10	0.7076	0.0385	0.6244	0.7757
22	73	1	15	0.6968	0.0394	0.6119	0.7666
23	57	0	10	0.6968	0.0394	0.6119	0.7666
24	47	0	41	0.6968	0.0394	0.6119	0.7666
25	6	0	3	0.6968	0.0394	0.6119	0.7666
26	3	0	1	0.6968	0.0394	0.6119	0.7666
27	2	0	1	0.6968	0.0394	0.6119	0.7666
37	1	0	1	0.6968	0.0394	0.6119	0.7666

Table 2B: Results of the univariable proportional hazards Cox regression model

Omnibus Tests of Model Coefficients

-2 Log Likelihood
401.960

Covariates	B	SE	Wald χ^2	df	Sig.	Exp(B)	LR χ^2	-2logL
Sex (Male)	-0.051	0.314	0.026	1	0.871	0.950	0.026	401.934
Age (18-34 years)			12.059	2	0.002 *		15.037	384.339
35-54 years	1.613	0.496	10.588	1	0.001	5.017		
>=55 years	1.706	0.517	10.890	1	0.001	5.505		
MRSUS (Single)			4.514	3	0.211		4.752	396.849
Married	0.819	0.389	4.435	1	0.035	2.268		
Separated/Divorced	0.624	0.558	1.250	1	0.264	1.866		
Widow/Widowed	0.771	0.667	1.336	1	0.248	2.162		
EDULABL (Illiterate)			1.513	3	0.679		1.535	400.393
Read and Write	-0.127	0.527	0.058	1	0.809	0.880		
Secondary	0.125	0.472	0.070	1	0.791	1.133		
Tertiary and above	-0.305	0.527	0.480	1	0.488	0.694		
EMPSUS (Employed)			1.164	3	0.762		1.175	400.813
Own Business	0.355	0.517	0.475	1	0.491	1.427		
Day laborer	0.385	0.671	0.330	1	0.566	1.470		
Unemployed	0.036	0.506	0.005	1	0.943	1.037		
Religion (Orthodox)			0.013	2	0.993		0.732	400.519
Muslim	10.024	177.105	0.003	1	0.995	22552.86		
Protestant	10.084	177.106	0.003	1	0.955	23949.16		
DELAY (>=1 month)	-1.047	0.315	11.057	1	0.001*	0.351	12.095	390.662
NODRUG (INH &RIF only)	1.520	0.364	17.462	1	<0.0001*	4.572	20.985	380.879
Smokstatus (Smokers)	-1.967	0.329	35.689	1	<0.0001*	0.140	48.330	363.796
Alcoholuse (Yes)	-2.206	0.599	13.544	1	0.0002*	0.110	19.974	376.999
(No Complication)			23.625	4	0.0001*		28.938	378.356
Pneumonia	1.831	0.423	18.704	1	0.000	6.239		
Pneumothorax	0.997	0.458	4.742	1	0.029	2.710		
Hemoptysis	1.223	0.517	5.586	1	0.018	3.397		
Cor pulmonal	1.645	0.484	11.571	1	0.001	5.181		
Cat (Previously treat)	1.887	0.313	36.388	1	0.000*	6.601	47.915	368.893
HIV (Positive)	-1.976	0.326	36.675	1	<0.0001*	0.139	49.487	372.953
Chronic (no chronic)			38.914	3	<0.0001*		54.840	369.589
Diabetes mellitus	1.085	0.419	6.713	1	0.001	2.958		

Myocardial infraction	2.515	0.486	26.770	1	0.000	12.362		
Asthma	1.987	0.421	22.274	1	0.000	7.292		
SMEAR (Positive)	0.403	0.362	1.239	1	0.266	1.497	1.256	400.810.

- *The reference categories are those indicated in brackets * the covariate is significant at the 0.05 level.*

Table 3B: Result of the multivariable proportional hazard regression model containing the variables significant in the univariabe proportional hazard model.

Omnibus Tests of Model Coefficients

-2 Log Likelihood
285.062

Covariates	B	SE	Wald X^2	df	Sig.	Exp(B)
DELAY(>=1 month)	-1.098	0.456	5.809	1	0.016	0.333
NODRUG (INH &RIF only)	-1.264	0.423	8.947	1	0.003	0.282
Smokstatus (smokers)	-0.685	0.392	3.043	1	0.081	0.505
Alcoholuse (Yes)	-1.462	0.630	5.383	1	0.020	0.232
(No Complication) Pneumonia Pneumothorax Hemoptysis Cor pulmonal	 0.955 -0.254 0.463 1.641	 0.516 0.554 0.599 0.567	10.384 3.427 0.211 0.596 8.391	4 1 1 1 1	0.034 0.064 0.646 0.440 0.004	 2.598 0.775 1.588 5.161
HIV (Positive)	-1.602	0.392	16.726	1	0.000	0.201
Chronic (no chronic) Diabetes mellitus Myocardial infraction Asthma	 1.015 2.309 1.434	 0.545 0.594 0.501	18.543 3.470 15.125 8.204	3 1 1 1	0.000 0.062 0.000 0.004	 2.759 10.062 4.196

- The reference categories are those indicated in brackets.

Table 4B: Result of Wald statistic P-values when possible interactions terms included in reduced model one at a time

Interaction between covariates		DF	Wald	P-Value
Therapeutic delay	Number of Drugs	1	0.605	0.437
	Alcohol use	1	0.337	0.561
	Clinical complication	4	0.995	0.911
	HIV Co-infection	1	0.291	0.590

	Chronic disease	3	0.640	0.887
Number of Drugs	Alcohol use	1	0.489	0.484
	Clinical complication	4	1.372	0.849
	HIV Co-infection	1	0.589	0.443
	Chronic disease	3	0.824	0.844
Alcohol use	Clinical complication	4	0.907	0.924
	HIV Co-infection	1	0.041	0.839
	Chronic disease	3	0.272	0.965
Clinical complication	HIV Co-infection	4	5.722	0.221
	Chronic disease	12	8.270	0.764
HIV Co-infection	Chronic disease	3	1.699	0.637

Table 5B: Result of the Exponential regression model with corresponding AIC and BIC values

Exponential regression -- log relative-hazard form

No. of subjects = 146 Number of obs = 146 No. of failures = 42

Time at risk = 2655 LR chi^2 (15) = 110.04

Log likelihood = -77.325422 Prob > chi2 = 0.0000

_t	Coef.	Std. Err.	Z	P>|z|	[95% Conf. Interval]	
					Lower	Upper
Sex	0.0452588	0.4166544	0.12	0.908	-0.7684673	0.8647879
Age	0.6768404	0.2812544	2.41	0.016	0.1255919	1.228089
MRSUS	-0.2481689	0.2461084	-1.01	0.313	-0.7305325	0.2341948
EDULABL	-0.2191629	0.220709	-0.99	0.321	-0.6517445	0.2134188
EMPSUS	-0.0376534	0.1360608	-0.28	0.782	-0.3043277	0.2290208
Religion	-0.2032551	0.6883346	-0.30	0.768	-1.552366	1.145856
DELAY	-1.060755	0.402847	-2.63	0.008	-1.85032	-0.2711891
NODRUG	-0.9018496	0.3901495	-2.31	0.021	-1.666529	-0.1371706
Smokstatus	-0.724259	0.3992386	-1.81	0.070	-1.506752	0.0582344
Alcoholuse	-1.444988	0.6567524	-2.20	0.028	-2.732199	-0.1577769
Clnccomplicn	0.2377893	0.1342939	1.77	0.077	-0.025422	0.5010006
MDRCAT	0.5181894	0.417035	1.24	0.214	-0.2991842	1.335563
HIV	-1.131205	0.438505	-2.58	0.010	-1.990659	-0.2717514
CHRONIC	0.471997	0.1572761	3.00	0.003	0.1637415	0.7802524

SMEAR	0.3752426	0.537844	0.70	0.485	-0.6789123	1.429398
_cons	-2.563492	0.9821159	-2.61	0.009	-4.488404	-0.638580

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	146	-132.3478	-77.32542	16	186.6508	234.3886

Note: N=Obs used in calculating BIC; see [R] BIC note

Table 6B: Result of the Weibull regression model with corresponding AIC and BIC values

Weibull regression -- log relative-hazard form

No. of subjects = 146 Number of obs = 146 No. of failures = 42

Time at risk = 2655 LR chi^2 (15) = 119.39

Log likelihood = -72.650423 Prob > chi2 = 0.0000

_t	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
					Lower	Upper
Sex	-0.032851	0.4370747	-0.08	0.940	-0.8895017	0.8237998
Age	0.8000984	0.2970161	2.69	0.007	0.2179576	1.382239
MRSUS	-0.294247	0.2521145	-1.17	0.243	-0.7883824	0.1998883
EDULABL	-0.2265574	0.2338126	-0.97	0.333	-0.6848217	0.2317068
EMPSUS	-0.0144881	0.1405208	-0.10	0.918	-0.2899038	0.2609276
Religion	-0.3911726	0.7305728	-0.54	0.592	-1.823069	1.040724
DELAY	-1.225845	0.4204272	-2.92	0.004	-2.049867	-0.4018227
NODRUG	-1.007244	0.3955844	-2.55	0.011	-1.782576	-0.2319133
Smokstatus	-0.7865751	0.4104426	-1.92	0.055	-1.591028	0.0178775
Alcoholuse	-1.395812	0.6557236	-2.13	0.033	-2.681006	-.1106173
Clnccomplicn	0.2576345	0.1382996	1.86	0.062	-0.0134278	0.5286968
MDRCAT	0.7362727	0.4412978	1.67	0.095	-0.1286552	1.601201
HIV	-1.256401	0.4674504	-2.69	0.007	-2.172587	-0.3402154
CHRONIC	0.5414255	0.1612987	3.36	0.001	0.225286	0.8575651
SMEAR	0.2253239	0.5596175	0.40	0.687	-0.8715061	1.322154
_cons	-4.126206	1.174263	-3.51	0.000	-6.427719	-1.824693
/ln_p	0.4502536	0.1350796	3.33	0.001	0.1855025	0.7150048

P	1.56871	0.2119007			1.203823	2.044196
1/p	0.6374664	0.0861087			0.4891898	0.830686

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	146	-132.345	-72.65042	17	179.3008	230.0222

Note: N= Obs used in calculating BIC; see [R] BIC note

Table 7B: results of the step wise variable selection of weibull regression analysis.

p = 0.0826 >= 0.0500 removing Smokstatus

Weibull regression -- log relative-hazard form

No. of subjects = 146 Number of obs = 146

No. of failures = 42

Time at risk = 2655

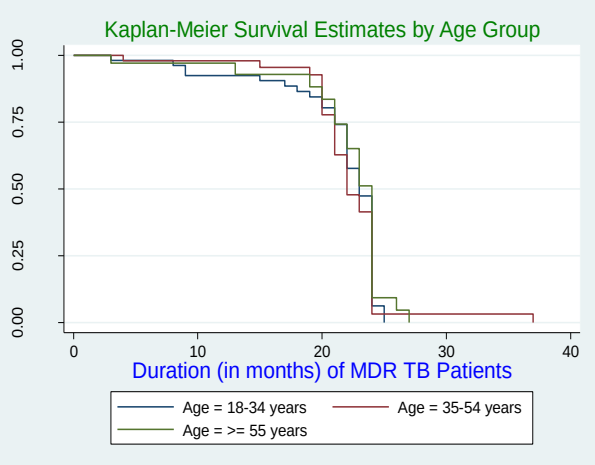
LR $chi^2(8)$ = 113.92

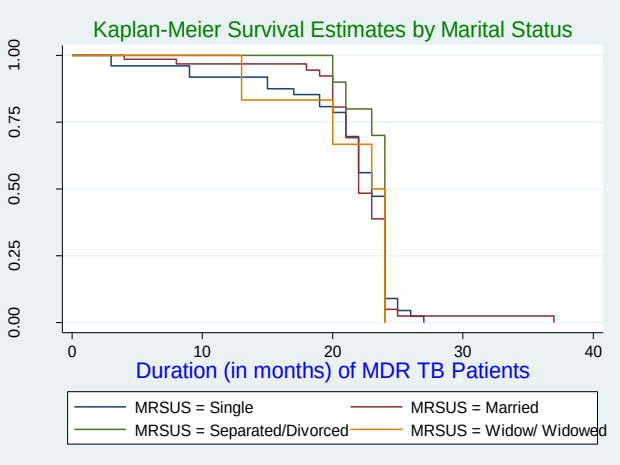
Log likelihood = -75.385077 Prob > chi2 = 0.0000

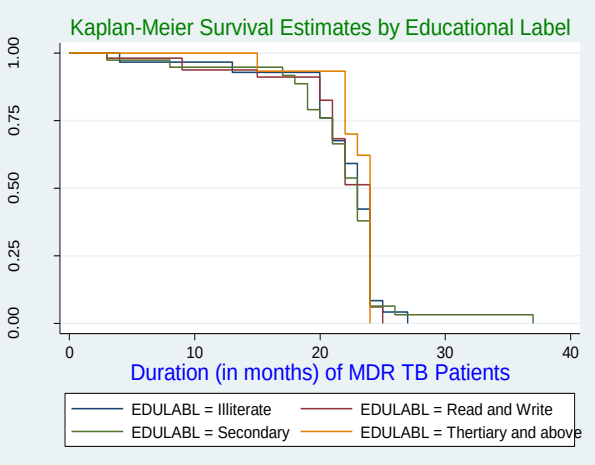
_t	Coef.	Std. Err.	z	P>z	[95% Conf. Interval]	
					Lower	upper
Age	0.748324	0.2422695	3.09	0.002	0.2734844	1.223164
DELAY	-1.219319	0.3421733	-3.56	0.000	-1.889967	-0.548672
NODRUG	-1.111739	0.3778737	-2.94	0.003	-1.852357	-0.3711196
CHRONIC	0.5587804	0.1339457	4.17	0.000	0.2962517	0.8213091
Alcoholuse	-1.46689	0.6164538	-2.38	0.017	-2.675117	-0.2586624
Clnccomplicn	0.3121017	0.1206824	2.59	0.010	0.0755685	0.5486349
MDRCAT	0.9551259	0.3541405	2.70	0.007	0.2610233	1.649228
HIV	-0.9944914	0.3669531	-2.71	0.007	-1.713706	-0.2752765
_cons	-5.265299	0.8774824	-6.00	0.000	-6.985133	-3.545465
ln_p	0.4247582	0.1320765	3.22	0.001	0.1658931	0.6836233
p	1.529221	0.201974			1.180447	1.981043
1/p	0.6539279	0.0863685			0.5047847	0.8471368

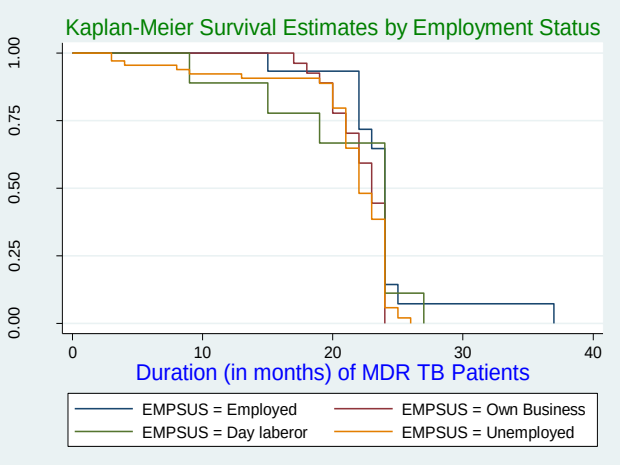
Appendix C

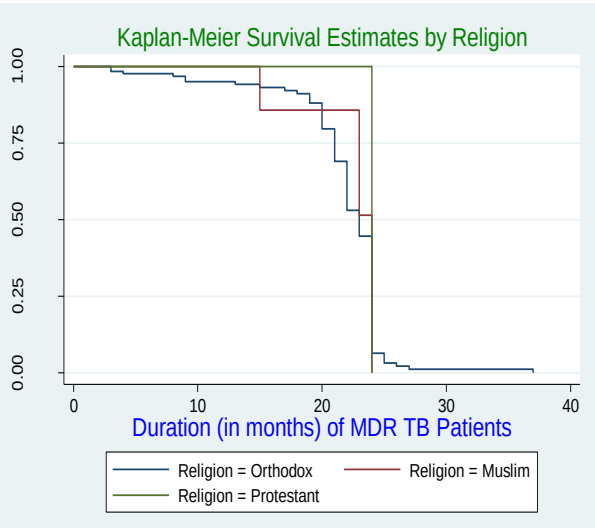
Figures 1C: Plots of Kaplan-Meier survivor functions, for different covariates

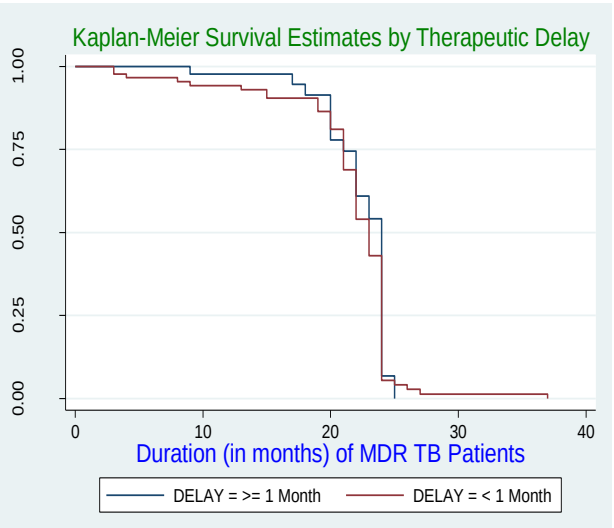


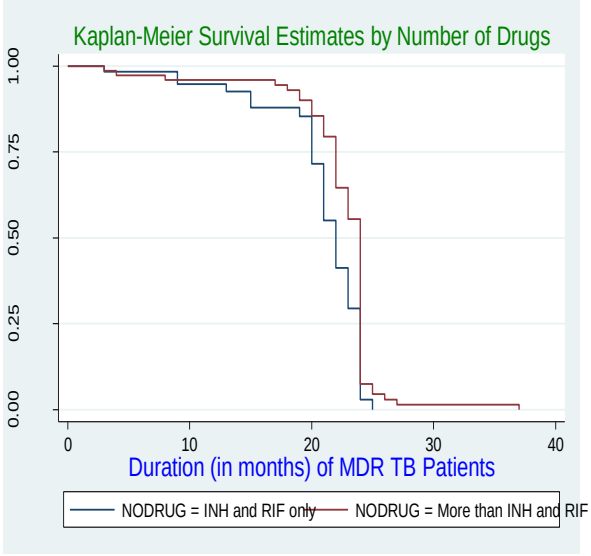


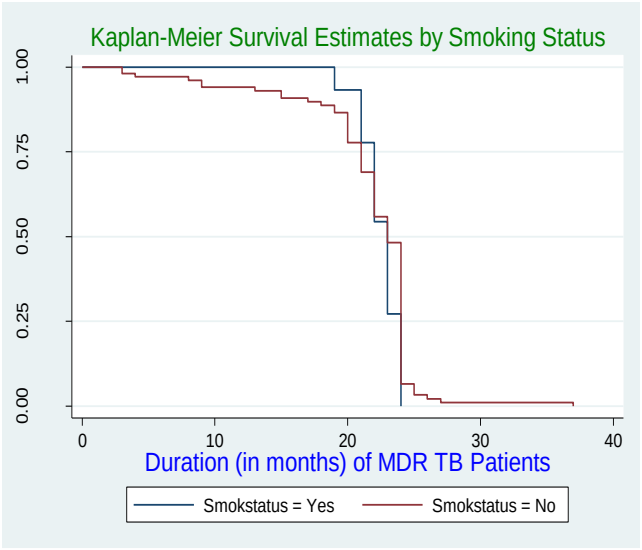


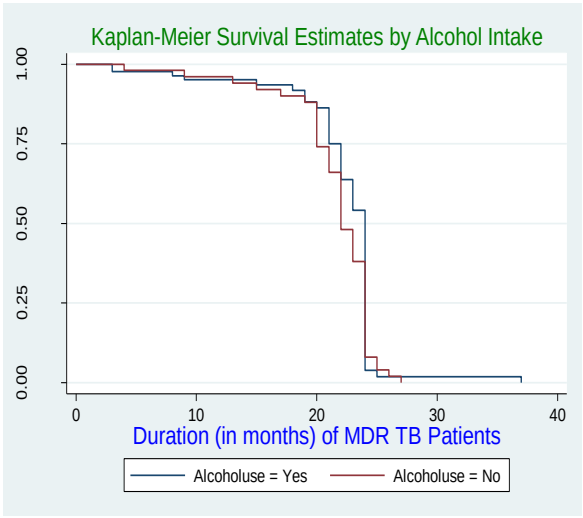


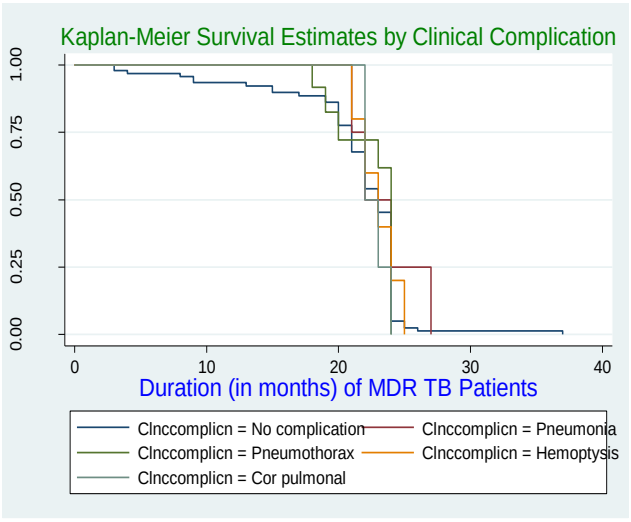


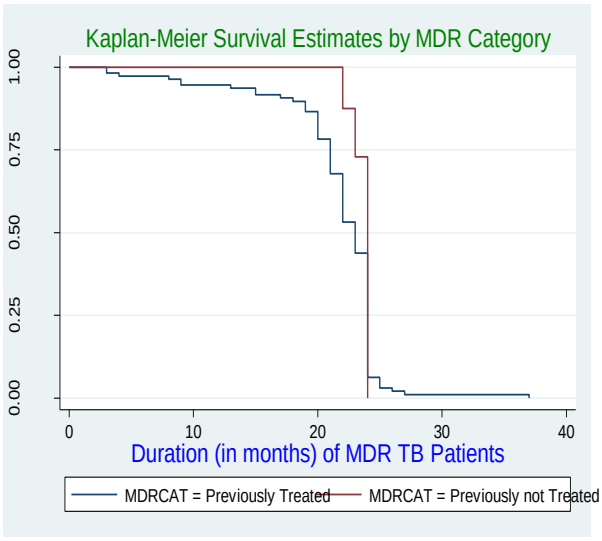


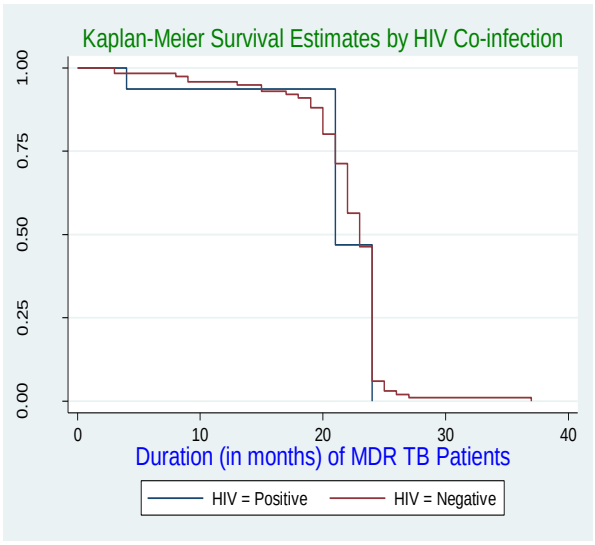

Kaplan-Meier Survival Estimates by Number of Drugs
0.00
0.25
0.50
0.75
1.00
0
10
20
30
40
Duration (in months) of MDR TB Patients
NODRUG = INH and RIF only
NODRUG = More than INH and RIF

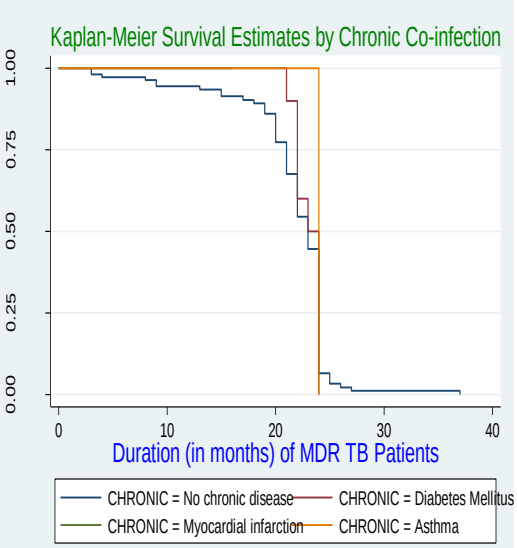

Kaplan-Meier Survival Estimates by Smoking Status
0.00
0.25
0.50
0.75
1.00
0
10
20
30
40
Duration (in months) of MDR TB Patients
Smokstatus = Yes
Smokstatus = No

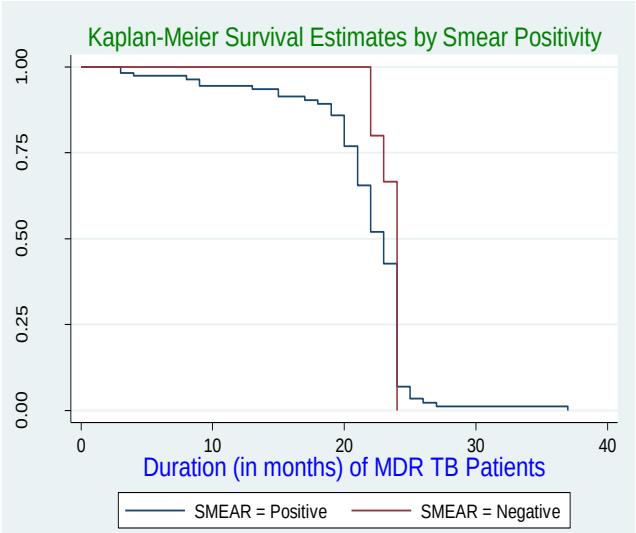

Kaplan-Meier Survival Estimates by Alcohol Intake
0.00
0.25
0.50
0.75
1.00
0
10
20
30
40
Duration (in months) of MDR TB Patients
Alcoholuse = Yes
Alcoholuse = No

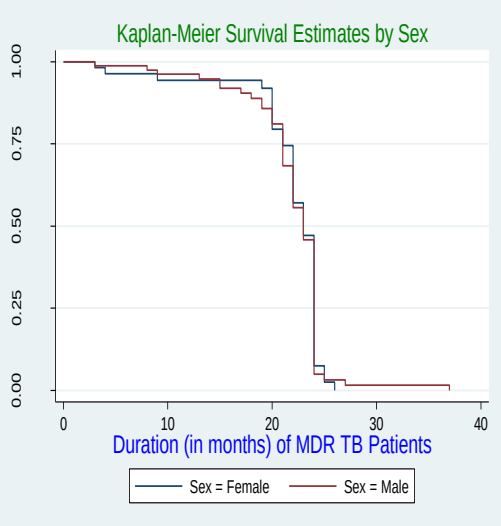

Kaplan-Meier Survival Estimates by Clinical Complication
0.00
0.25
0.50
0.75
1.00
0
10
20
30
40
Duration (in months) of MDR TB Patients
ClnccomplicN = No complication
ClnccomplicN = Pneumonia
ClnccomplicN = Pneumothorax
ClnccomplicN = Hemoptysis
ClnccomplicN = Cor pulmonal


Kaplan-Meier Survival Estimates by MDR Category
0.00
0.25
0.50
0.75
1.00
0
10
20
30
40
Duration (in months) of MDR TB Patients
MDRCAT = Previously Treated
MDRCAT = Previously not Treated


Kaplan-Meier Survival Estimates by HIV Co-infection
0.00
0.25
0.50
0.75
1.00
0
10
20
30
40
Duration (in months) of MDR TB Patients
HIV = Positive
HIV = Negative







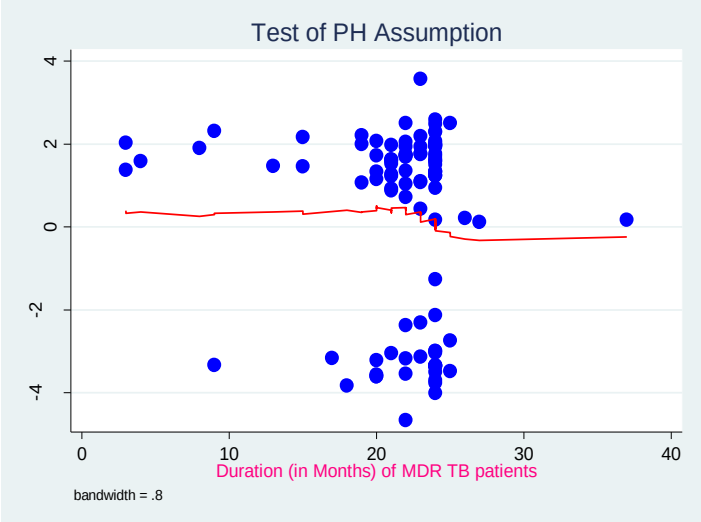
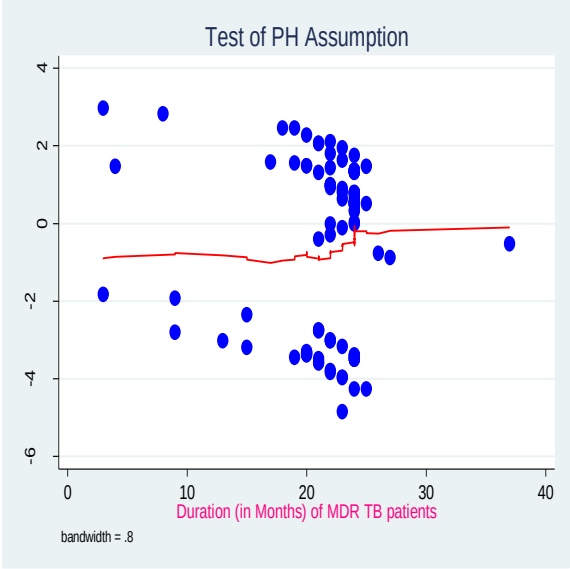


Figure 2C: Plots of the Scaled Schoenfeld residuals & their lowess smooth obtained from the final model for the covariate Therapeutic Delay *of the patient*

.



.

Figure 3C: Plots of the Scaled Schoenfeld residuals & their lowess smooth obtained from the final model for the covariate Number of drugs at initiation

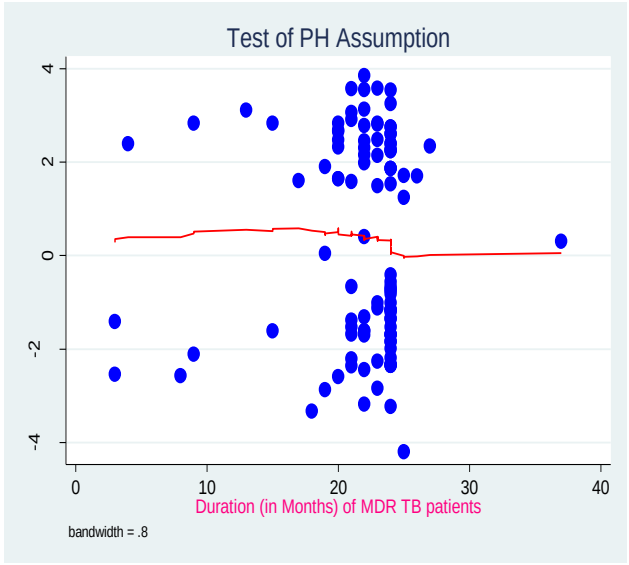


Figure 4C: Plots of the Scaled Schoenfeld residuals & their lowess smooth obtained from the final model for the covariate Alcohol intake of the patient.

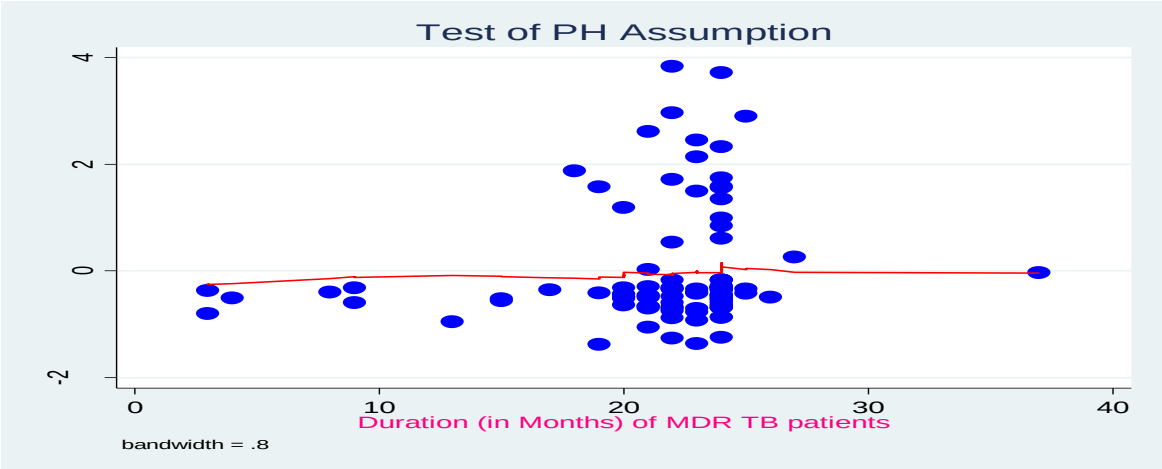


Figure 5C: Plots of the Scaled Schoenfeld residuals and their lowess smooth obtained from the final model for the covariate any clinical complication of the patient.

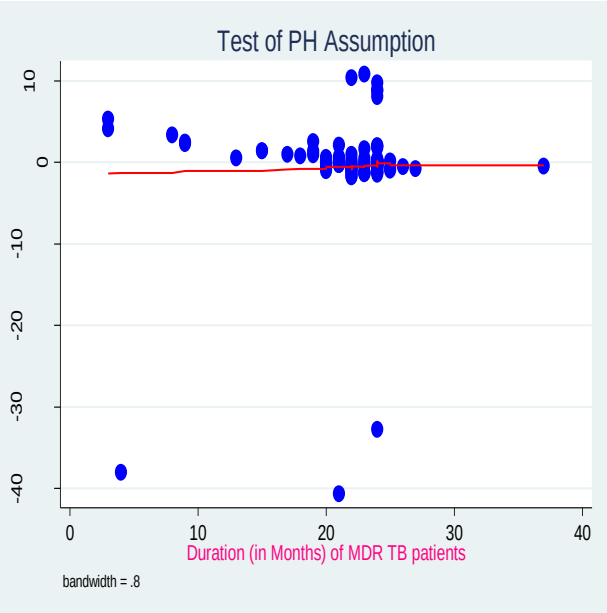


Figure 6C: Plots of the Scaled Schoenfeld residuals & their lowess smooth obtained from the final model for the covariate HIV co_infection of the patient

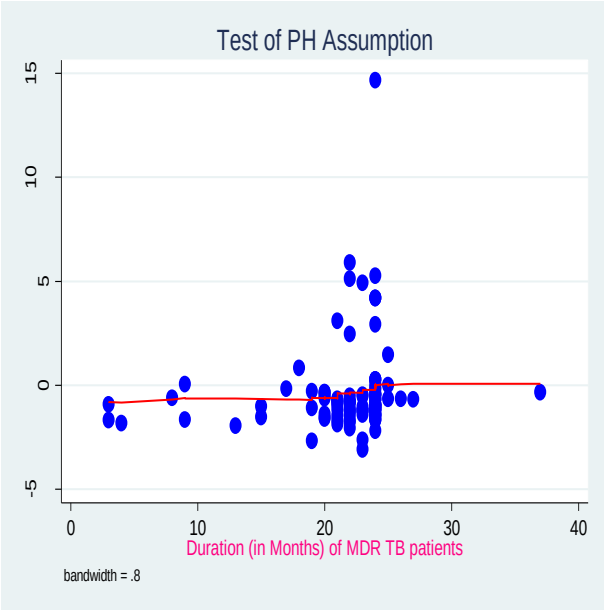


Figure 7C: Plots of the Scaled Schoenfeld residuals & their lowess smooth obtained from the final model for the covariate chronic co_infection of the patient.

Appendix D: STATA and SAS codes

STATA codes

```
global time  TIME
global event  STATUS

stset TIME, failure(STATUS)
stdescribe
stsum

label var  Sex "gender of the patient"
label define Sex 0"Male" 1"Female"
label value Sex Sex

          gen Age=.
          replace Age=0 if Age>18 & Age<=34
          replace Age=1 if Age>34 & Age<=54
          replace Age =2 if Age>54
          label var Age "Age of the patient"
          label define Age 0"18-34 years " 1"34-54 years " 2"55 & above"
          label value Age Age
          tab Age

label var MRSUS "marital status of the patient"
label define MRSUS 0"Single" 1"Married" 2"Separated/Divorced"
3"Widow/Widowed"
label value MRSUS MRSUS

label var EDULABL "Educational label of the patient "
label define EDULABL 0"Illiterate" 1"Read and Write" 2"Secondary"
3"Teritiary and above"
label value EDULABL EDULABL

label var EMPSUS "Employment status of the patient "
label define EMPSUS 0"Employed" 1"Own Business" 2"day laborer"
3"Unemployed"
label value EMPSUS EMPSUS

label var Religion " Religion of the patient "
label define Religion 0"Orthodox" 1"Muslim" 2"Protestant" 3"Others"
label value Religion Religion

label var DELAY "Therapeutic Delay"
label define DELAY 0">= 1 month" 1"< 1 month"
label value DELAY DELAY

label var NODRUG "Number of drugs at initiation"
label define NODRUG 0"INH and RIF only" 1"more than INH and RIF"
label value NODRUG NODRUG

label var Smokstatus " Smoking status"
label define Smokstatus 0"Yes 1"No"
```

```
label Smokstatus Smokstatus


label var Alcoholuse " Alcohol use"
label define Alcoholuse 0"Yes 1"No"
label Alcoholuse Alcoholuse

label var Clnccomplicn "any clinical complication"
label define Clnccomplicn 0"No complication" 1"Pneumonia" 2"Pneumothorax"
3"Hemoptysis" 4"Cor pulmonal"
label value Clnccomplicn  Clnccomplicn

label var MDRCAT " MDR Category"
label define MDRCAT 0"Previously treated"  1"Previously not Treated"
label MDRCAT MDRCAT


label var HIV " HIV co-infection"
label define HIV 0"Positive"  1"Negative"
label HIV HIV

label var CHRONIC "any clinical complication"
label define CHRONIC 0"No chronic disease" 1"Diabetes mellitus"
2"Myocardial infraction " 3"Asthma"
label value CHRONIC CHRONIC

label var SMEAR " Smear positivity"
label define SMEAR 0"Positive"  1"Negative"
label SMEAR SMEAR

streg Sex Age MRSUS EDULABL EMPSUS Religion DELAY NODRUG Smokstatus
Alcoholuse Clnccomplicn MDRCAT HIV CHRONIC SMEAR ,nohr dist(exponential)

streg i.Age DELAY  NODRUG   Alcoholuse  i.Clnccomplicn  MDRCAT  HIV i.CHRONIC, nohr
dist(weibull)

streg Sex Age MRSUS EDULABL EMPSUS Religion DELAY NODRUG Smokstatus
Alcoholuse Clnccomplicn MDRCAT HIV CHRONIC SMEAR ,nohr dist(weibull)

xi: stepwise,pr(0.05): streg  Age DELAY  NODRUG  Smokstatus  Alcoholuse  Clnccomplicn
MDRCAT  HIV CHRONIC,nohr dist(weibull)

streg Age DELAY  NODRUG   Alcoholuse  Clnccomplicn  MDRCAT  HIV CHRONIC,
nohr  dist(weibull)
scalar m1 = e(ll)
estimates store m1
streg  ,nohr dist(weibull)
scalar m2 = e(ll)
estimates store m2
lrtest m2 m1

estat ic: stata command for computing AIC and BIC after a model have been
already fitted
```

SAS codes

```
data MDR;
input DELAY  NODRUG  Smokstatus  Alcoholuse  Clnccomplicn  HIV   CHRONIC  TIME
STATUS;
datalines;

;
proc print;
run;
proc phreg data=MDR;
model TIME*STATUS(0)= DELAY  NODRUG  Smokstatus  Alcoholuse  Clnccomplicn  HIV
CHRONIC;
run;
proc phreg data=MDR;
title "test of proportionality assumption";
model TIME*STATUS(0)= DELAY  NODRUG  Smokstatus  Alcoholuse  Clnccomplicn  HIV
CHRONIC DELAYT NODRUGT SmokstatusT AlcoholuseT
ClnccomplicnT HIVT CHRONICT;
DELAYT = DELAY*log(TIME);
NODRUGT = NODRUG*log(TIME);
SmokstatusT = Smokstatus*log(TIME);
AlcoholuseT = Alcoholuse*log(TIME);
ClnccomplicnT = Clnccomplicn*log(TIME);
HIVT = HIV*log(TIME);
CHRONICT =  CHRONIC*log(TIME);
test_proportionality: test DELAY, NODRUG, Smokstatus, Alcoholuse, Clnccomplicn, HIV,
CHRONIC;
run;

data MDR;
input Sex Age  MRSUS  EDULABL  EMPSUS  Religion  DELAY  NODRUG  Smokstatus  Alcoholuse
Clnccomplicn  MDRCAT  HIV CHRONIC SMEAR  TIME  STATUS;
datalines;

;
proc print;
run;
proc lifereg data=MDR;
model TIME*STATUS(0) =  Age  DELAY  NODRUG  Smokstatus  Alcoholuse  Clnccomplicn
MDRCAT  HIV CHRONIC  / distribution=weibull;
run;

proc lifereg data=MDR noprint;
model TIME*STATUS(0) = Sex Age  MRSUS  EDULABL  EMPSUS  Religion  DELAY  NODRUG
Smokstatus  Alcoholuse  Clnccomplicn  MDRCAT  HIV CHRONIC SMEAR /
distribution=exponential;
output out= exp cdf=f;
run;
data exp1;
  set exp;
```

```
  cox = -log( 1-f );
run;
proc lifetest data=exp1 outsurv=surv_exp noprint;
TIME cox*STATUS(0);
run;
data surv_exp;
  set surv_exp;
  ls = -log(survival);
run;
goptions reset=all;
axis1 order=(0 to 5 by 1) minor=none label=('Exponential Reg Model Cum Hazard');
axis2 order=(0 to 5 by 1) minor=none label=( a=90 'Kaplan-Meier Cum Hazard');
symbol1 i=l1p  c= blue v=dot h=.4;
symbol2 i = join c = red l = 3;
proc gplot data=surv_exp;
  plot (ls cox)*cox / overlay haxis=axis1 vaxis= axis2;
run;
quit;

proc lifereg data=MDR noprint;
model TIME*STATUS(0) = Sex Age  MRSUS  EDULABL  EMPSUS  Religion  DELAY  NODRUG
Smokstatus  Alcoholuse  Clnccomplicn  MDRCAT  HIV CHRONIC SMEAR /
distribution=weibull;
output out= weibull cdf=f;
run;
data weibull1;
set weibull;
cox = -log( 1-f );
run;
proc lifetest data=weibull1 outsurv=surv_wei noprint;
 TIME cox*STATUS(0);
run;
data surv_wei;
  set surv_wei;
  ls = -log(survival);
run;
goptions reset=all;
axis1 order=(0 to 5 by 1) minor=none label=('Weibull Reg Model Cum Hazard');
axis2 order=(0 to 5 by 1) minor=none label=( a=90 'Kaplan-Meier Cum Hazard');
symbol1 i=l1p  c= blue v=dot h=.4;
symbol2 i = join c = red l = 3;
proc gplot data=surv_wei;
  plot (ls cox)*cox / overlay haxis=axis1 vaxis= axis2;
run;
quit;
```